A Nomadic Subtelomeric Disease Resistance Gene Cluster in Common Bean[1][W]

Perrine David, Nicolas W.G. Chen, Andrea Pedrosa-Harand, Vincent Thareau, Mireille Sévignac, Steven B. Cannon, Daniel Debouck, Thierry Langin, and Valérie Geffroy*

Institut de Biotechnologie des Plantes, UMR-CNRS 8618, Université Paris-Sud, 91405 Orsay cedex, France (P.D., N.W.G.C., V.T., M.S., T.L., V.G.); Unité Mixte de Recherche de Génétique Végétale, Institut National de la Recherche Agronomique, 91190 Gif-sur-Yvette, France (V.G.); Laboratório de Citogenética Vegetal, Departamento de Botânica-Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Recife-Pernambuco 50670–420, Brazil (A.P.-H.); United States Department of Agriculture-Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, Iowa 50011 (S.B.C.); and Genetic Resources Unit, Centro Internacional de Agricultura Tropical, AA 6713 Cali, Colombia (D.D.)

The B4 resistance (*R*) gene cluster is one of the largest clusters known in common bean (*Phaseolus vulgaris* [*Pv*]). It is located in a peculiar genomic environment in the subtelomeric region of the short arm of chromosome 4, adjacent to two heterochromatic blocks (knobs). We sequenced 650 kb spanning this locus and annotated 97 genes, 26 of which correspond to Coiled-Coil-Nucleotide-Binding-Site-Leucine-Rich-Repeat (CNL). Conserved microsynteny was observed between the *Pv* B4 locus and corresponding regions of *Medicago truncatula* and *Lotus japonicus* in chromosomes *Mt6* and *Lj2*, respectively. The notable exception was the CNL sequences, which were completely absent in these regions. The origin of the *Pv* B4-CNL sequences was investigated through phylogenetic analysis, which reveals that, in the *Pv* genome, paralogous CNL genes are shared among nonhomologous chromosomes (4 and 11). Together, our results suggest that *Pv* B4-CNL was derived from CNL sequences from another cluster, the Co-2 cluster, through an ectopic recombination event. Integration of the soybean (*Glycine max*) genome data enables us to date more precisely this event and also to infer that a single CNL moved from the Co-2 to the B4 cluster. Moreover, we identified a new 528-bp satellite repeat, referred to as *khipu*, specific to the *Phaseolus* genus, present both between B4-CNL sequences and in the two knobs identified at the B4 *R* gene cluster. The *khipu* repeat is present on most chromosomal termini, indicating the existence of frequent ectopic recombination events in *Pv* subtelomeric regions. Our results highlight the importance of ectopic recombination in *R* gene evolution.

In the human genome, extensive cytogenetic and sequence analyses have revealed that subtelomeres are hot spots of interchromosomal recombination and segmental duplications (Linardopoulou et al., 2005). This peculiar dynamic activity of subtelomeres has been reported in such diverse organisms as yeast and the malaria parasite *Plasmodium* (Louis, 1995; Freitas-Junior et al., 2000, 2005). As expected for a plastic region of the genome subject to reshuffling through recombination events, subtelomeres exhibit unusually high levels of within-species structural and nucleotide polymorphism (Mefford and Trask, 2002). In plants, this plasticity of subtelomeres has not been identified in Arabidopsis (*Arabidopsis thaliana*; Heacock et al., 2004; Kuo et al., 2006) and, to our knowledge, has not yet been investigated at a large scale for other plant species with full genome sequences available. Regarding Arabidopsis, the apparent lack of high subtelomeric recombination may reflect its small and simple subtelomeres, mirroring its small genome size and relative paucity of repetitive sequences (Heacock et al., 2004; Kuo et al., 2006).

Repetitive sequences, such as satellite DNA and retroelements, constitute an important fraction of every eukaryotic genome and therefore constitute the environment in which genes are expressed. Satellite DNA can be defined as highly reiterated noncoding DNA sequences, organized as long arrays of head-to-tail linked repeats of 150- to 180-bp or 300- to 360-bp monomers located in the constitutive heterochromatin (Plohl et al., 2008). Despite their ubiquity in eukaryotic genomes, little is known about the mechanisms that allow these elements to accumulate. Early hypotheses considered them to be nonfunctional "selfish" or "junk" DNA segments that increase or decrease their frequency without any advantage or disadvantage for an organism (Ohno, 1972; Orgel and Crick, 1980). However, identification of satellite DNA at structurally

[1] This work was supported by the Institut National de la Recherche Agronomique, the Centre National de la Recherche Scientifique, and a sequencing grant from the Ministère de la Recherche, Centre National de la Recherche Scientifique to V.G.

* Corresponding author; e-mail valerie.geffroy@u-psud.fr.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Valérie Geffroy (valerie.geffroy@u-psud.fr).

[W] The online version of this article contains Web-only data.

www.plantphysiol.org/cgi/doi/10.1104/pp.109.142109

important parts of chromosomes, such as centromeres, has suggested functional roles of satellite DNA (Ma and Jackson, 2006; Kawabe and Charlesworth, 2007). Satellite DNA can also be localized in knobs, which are cytologically visible regions of highly condensed chromatin (heterochromatin) that are distinct from pericentromeric regions in pachytene chromosomes (Fransz et al., 2000; Gaut et al., 2007; Lamb et al., 2007).

The survival of most organisms depends on the presence of specific genetic systems that maintain diversity in order to respond to changing environments. Plants, like animals, are continually challenged by a large array of pathogens. To perceive and counter pathogen attack, plants have evolved disease resistance (*R*) genes. The largest class of *R* genes encodes proteins containing a central Nucleotide-Binding Site (NBS) domain, a C-terminal Leucine-Rich Repeat (LRR) domain, and a variable N-terminal domain. These R proteins detect the presence of disease-causing bacteria, oomycetes, fungi, nematodes, insects, and viruses by sensing either specific pathogen effector molecules produced during the infection process or key molecules in the plant cell that may be attacked by pathogen effectors (Dangl and McDowell, 2006). The evolution of new *R* genes serves to counteract the evolution of novel virulence factors from the pathogens (McDowell and Simon, 2008). Among this prevalent class of *R* gene, two subclasses, corresponding to two ancient lineages (Bai et al., 2002; Meyers et al., 2003; Ameline-Torregrosa et al., 2008), have been identified based on the N-terminal domain of the R protein: the Coiled-Coil (CC)-NBS-LRR (CNL) and the Toll-Interleukin receptor (TIR)-NBS-LRR (TNL). Genome studies have demonstrated that NBS-LRR (NL) sequences are abundant in any plant genome. For example, annotation of the Arabidopsis, rice (*Oryza sativa*), poplar (*Populus trichocarpa*), *Medicago truncatula* (*Mt*), grape (*Vitis vinifera*), *Lotus japonicus* (*Lj*), and papaya (*Carica papaya*) genomes identified at least 149, 480, 317, 333, 233, 229, and 55 genes encoding NL proteins, respectively (Bai et al., 2002; Meyers et al., 2003; Zhou et al., 2004; Tuskan et al., 2006; Velasco et al., 2007; Ameline-Torregrosa et al., 2008; Kohler et al., 2008; Ming et al., 2008; Sato et al., 2008). NL sequences are often located at complex loci (Smith et al., 2004), as exemplified by Arabidopsis, where two-thirds of them are organized in tightly linked clusters (Meyers et al., 2003; Leister, 2004; McDowell and Simon, 2006). Evolution of NL sequences in the Arabidopsis genome has been investigated according to their phylogenetic positions and physical locations. Although tandem duplications explain the origin of a large fraction of NLs, it seems that ectopic recombination has also played a role in Arabidopsis NL evolution, since mixed clusters comprising evolutionarily distant NL exist. Ectopic recombination is also evident when phylogenetically close *R* genes are physically dispersed on different chromosomes (Leister, 2004; McDowell and Simon, 2006). These results confirm pioneer macrosynteny studies between related monocot species suggesting the existence of NL movement in plant genomes. Indeed, extensive loss of collinearity between NL sequences between rice and barley (*Hordeum vulgare*), which diverged 50 million years ago (Mya), has suggested rapid reorganization of NL sequences (Leister et al., 1998; Leister, 2004). However, our knowledge of the molecular evolution of *R* genes remains limited due to the still small number of complete plant genome sequences available to date. Detailed comparative study across taxa at different evolutionary distances is needed to see how *R* gene clusters evolve at various time scales.

Legumes (Fabaceae) constitute the third largest family of flowering plants and represent the second most important family of agronomically important plants after Poaceae (Graham and Vance, 2003). As a result of recent sequencing efforts, legumes are one of the few plant families with extensive genome sequences in different species, since the soybean (*Glycine max* [*Gm*]) genome sequence is complete (http://www.phytozome.net/soybean.php) and both *Mt* and *Lj* genome sequences are nearly complete (Young et al., 2005; Sato et al., 2008). Consequently, the legume family is extremely well adapted for comparative phylogenomic approaches, in which phylogenetic inference is combined with structural genomic analyses (Ammiraju et al., 2008). Common bean (*Phaseolus vulgaris* [*Pv*]) is the most important grain legume for direct human consumption (Broughton et al., 2003). *Pv* is a selfing species and has a small diploid genome (2n = 22) of 588 Mb (Bennett and Leitch, 1995). Conservation of genome macrostructure (macrosynteny) has been reported between several legumes, including common bean and the two model legume species *Mt* and *Lj* genomes (Zhu et al., 2005; Hougaard et al., 2008). However, the extent of gene order conservation at the DNA sequence level has not yet been evaluated within orthologous chromosome segments between *Pv* and the two model legume species.

In the genome of common bean, many disease *R* genes are clustered at complex loci located at the ends (rather than the centers) of linkage groups (LGs; Vallejos et al., 2006; Geffroy et al., 2008). For example, *Colletotrichum lindemuthianum* Co-2 *R* specificity maps at one end of LG B11 (Adam-Blondon et al., 1994). Molecular analysis has revealed that this locus consists of a tandem array of CNL sequences (Geffroy et al., 1998; Creusot et al., 1999). Another CNL-rich region has been identified at the end of LG B4 in the vicinity of *R* specificities and *R* quantitative trait loci against a large selection of pathogens, including *C. lindemuthianum*, *Uromyces appendiculatus*, and the bacterium *Pseudomonas syringae* (Geffroy et al., 1998, 1999; Miklas et al., 2006). Recently, fluorescence in situ hybridization (FISH) analysis revealed that this complex *R* cluster is located in the subtelomeric region of the short arm of chromosome 4 and includes two knobs (Geffroy et al., 2009). In a sequencing effort focused on CNL sequences, we have previously identified 17 CNL sequences of the B4 locus (referred to as B4-CNL) from

Pv genotype BAT93 (Ferrier Cana et al., 2003, 2005; Geffroy et al., 2009). In the BAT93 genotype, these B4-CNL sequences are located on both sides of the subterminal knob (Geffroy et al., 2009).

To investigate the organization and the evolutionary origin of the subtelomeric B4 *R* gene cluster, we have sequenced approximately 650 kb of the *Pv* B4 *R* gene cluster, revealing that, in genotype BAT93, CNL are spread out in four subclusters, separated by non-CNL-encoding genes. This *Pv* sequence was then compared gene by gene with the sequenced portions of the three sequenced legume genomes, *Mt*, *Lj*, and *Gm*. Conserved microsynteny (conservation of local gene repertoire, order, and orientation) was observed, except for the CNL sequences, which appear to be completely absent in the corresponding regions of *Mt* and *Lj*. In this study, by combining genomics, phylogenetic, and cytogenetic approaches, we provide evidence that ectopic recombination in subtelomeric regions between nonhomologous chromosomes (4 and 11), involving a single CNL, gave rise to the *Pv* B4 *R* gene cluster. Chromosomal distribution of a new satellite DNA tandem repeat, referred to as *khipu*, suggests that ectopic recombination events in subtelomeric regions of bean nonhomologous chromosomes are frequent. Our results highlight the importance of ectopic recombination as an important evolutionary mechanism for the evolution of disease resistance genes.

RESULTS

Sequencing, Annotation, and Analysis of the NBS-LRR Sequences of the B4 *R* Gene Cluster

The screening of the BAT93 bacterial artificial chromosome (BAC) clone library with the *PRLJ1* NBS probe specific to the B4 *R* gene cluster (Geffroy et al., 1999, 2009) led to the identification of 73 positive BAC clones organized in six contigs (David et al., 2008). We sequenced six BAC clones representing the minimum tilling path, providing 647,335 bp of nonoverlapping sequence split up into three contigs: the two nonoverlapping BAC clones 48-B10 (124.8 kb) and FZ-E9 (122.2 kb) and a 410.2-kb continuous region covered by BAC clones 94-L19, FY-N24, 254-G15, and 75-H11. BAC clones 48-B10 and FZ-E9 were mapped at genetic positions (GP) 1 and 4, respectively, while the 410-kb contig spans GP 2 and 3, separated by 0.7 centimorgan (cM). The genetic distances between these three contigs are shown in Figure 1A. The average GC content of the approximately 650 kb is 33.07%. Annotation showed a total of 79 genes or pseudogenes with an average density of one gene per 8.1 kb (Fig. 1A; Supplemental Tables S1 and S2). Sixteen transposable element (TE) sequences or TE fragments have been identified (Supplemental Table S3). Most of them (13) fell into the class I TE: 11 are LTR retroelements and two are non-LTR retroelements. Among the LTR retroelements, six are Ty1/*Copia*-like, one is Ty3/*Gypsy*-like, and four are unclassified retroelements. The class II TE is the smallest set, with three CACTA-type elements, all identified in the 48-B10 BAC clone.

A total of 26 CNL-encoding genes were identified in the approximately 650-kb sequence. Among the 26 CNL, nine contain stop codons and/or frameshifts (referred to as "truncated" genes or "pseudogenes"), while the other 17 present uninterrupted open reading frames. These 26 CNL are not uniformly distributed but are organized in four subclusters. Subclusters A, C, and D map to GP 1, 3, and 4 and contain five, 11, and three CNL, respectively; subcluster B, spanning GP 2 and 3, contains seven CNL. Within subclusters A, B, C, and D, CNL are separated on average by 15, 8.3, 5, and 6.6 kb, respectively, and are predicted to be in the same orientation within a given subcluster (Fig. 1A; Supplemental Fig. S1). For example, all the CNL in subcluster C are oriented with predicted transcription from centromere to telomere, whereas all the CNL in subcluster B have the opposite orientation compared with CNL from subcluster C. The B4 *R* gene cluster is composed of homogeneous CNL sequences, since all 26 BAT93 CNL sequences show over 75% nucleotide identity with an average value of 87% (Supplemental Table S4). Positive selection assessment conducted on this comprehensive B4-CNL data set confirmed previous results obtained on a subset of B4-CNL (Geffroy et al., 2009). In particular, positively selected residues were identified mainly on *x* residues predicted to be solvent exposed in the LRR domain (Supplemental Fig. S2).

In order to infer phylogenetic relationships among the BAT93 *Pv* B4-CNL sequences, we constructed two phylogenetic trees using the PhyML and MrBayes programs. No major differences were found between these two trees; thus, Figure 2 presents only the tree resulting from the MrBayes analysis. Most of the phylogenetically close CNL sequences, such as CNL-B9 and CNL-B10*, map to adjacent physical positions, while some others map to nonadjacent physical positions, regardless of the phylogenetic method employed. For example, CNL-B22 (subcluster D) and CNL-B15 (subcluster B) are closely related in the phylogenetic tree, while these sequences are separated by more than 350 kb. Likewise, CNL-B17, CNL-B1*, and CNL-B6* are phylogenetically closely related, while CNL-B17 (subcluster B) is approximately 269 kb and approximately 302 kb away from CNL-B6* (subcluster C) and CNL-B1* (subcluster C), respectively. Together, these patterns suggest that *Pv* B4-CNL sequences were amplified at the B4 locus mostly by local tandem duplication within a given subcluster and that other types of genetic exchange occurred between and/or among CNL of the four subclusters.

The *Pv* B4 *R* Gene Cluster and Homologous Regions in *Lj*, *Mt*, and *Gm* Are Highly Collinear Except for the CNL Sequences

Functional DNA sequences are likely to be conserved between species and therefore are an important

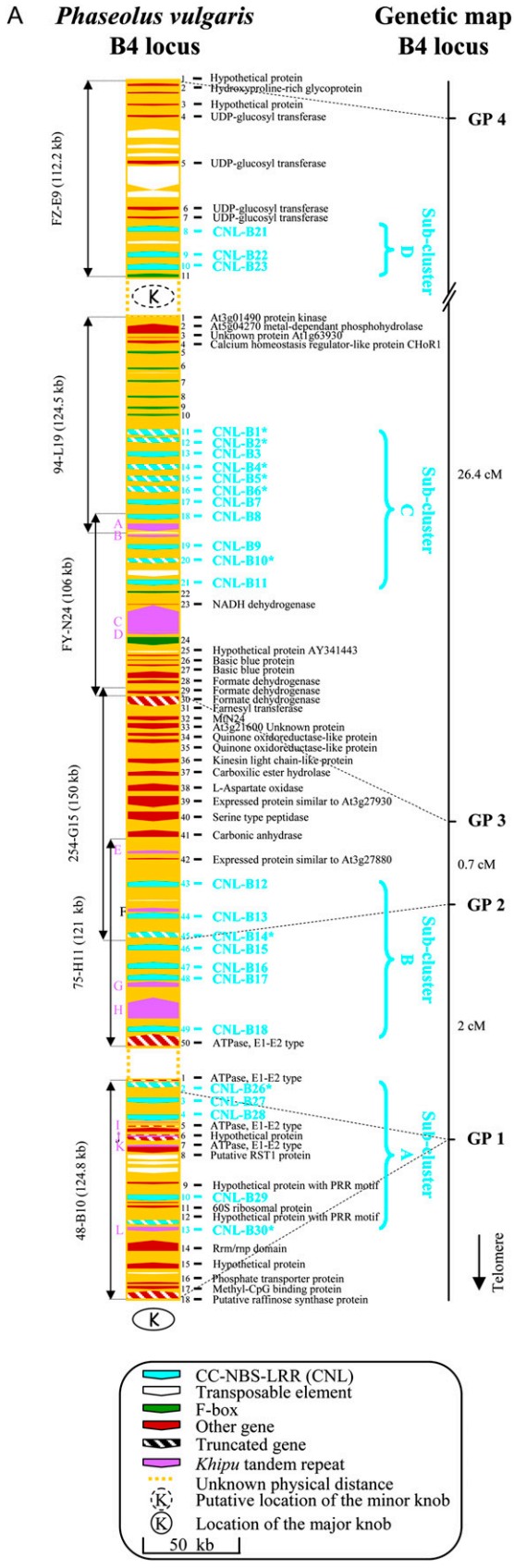

Figure 1. (*Figure continues on following page.*)

tool in comparative structural genomics (Bennetzen and Ma, 2003). Conceptual translations of DNA sequences were used for the comparisons because they provide a more sensitive test of homology between evolutionarily widely separated species than do nucleotide sequence comparisons. To detect syntenic regions, we used TBLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi) analysis on BAC clone sequences from the three sequenced or nearly completely sequenced genome species *Mt*, *Lj*, and *Gm*. Due to the nature of genes present in *Pv* BAC clones 48-B10 and FZ-E9 (high-copy genes), we could not conduct synteny analysis for these two *Pv* BAC clones. Synteny analysis was thus restricted to the *Pv* B4 410-kb contig containing CNL subclusters B and C (Fig. 1B; Supplemental Fig. S3).

In the *Lj* genome, a region syntenic to the *Pv* B4 locus was identified in a 360-kb contig (CM0201; accession no. DF093324) located on chromosome *Lj*2 (genetic position between 20.5 and 21.7 cM). In the *Pv* B4 410-kb contig, 17 out of the 25 predicted genes corresponding to low-copy genes in the Arabidopsis genome based on BLASTP analysis (in red in Fig. 1) have homologs in *Lj*, mostly conserved in order and orientation. Extensive collinearity is obvious between the *Pv* B4 locus and *Lj* contig CM0201, with the exception of CNL. Indeed, CNL from *Pv* subclusters B and C are completely absent in the *Lj* syntenic region.

Considering the same set of 25 *Pv* B4 low-copy genes, collinearity was observed between *Pv* B4 genes 36 to 50 and two *Mt* overlapping BAC clones (AC159962 and AC124951), located at the end of the short arm of chromosome *Mt6* (Fig. 1B). No *Mt* BAC clones overlapping with *Mt* AC159962 and AC124951 have yet been sequenced. All eight *Pv* B4 low-copy genes 36 to 42 and 50 have homologs in the *Mt* corresponding region, mostly conserved in order and orientation. Similar to what was observed with *Lj*, collinearity was strictly restricted to low-copy genes, and CNL from subcluster B are completely absent in the *Mt* syntenic region. Inspection of *Mt* and *Lj* genomic sequences revealed that no CNL-encoding genes are present in areas surrounding the identified *Pv* B4 locus syntenic regions.

After *Pv* and *Gm* diverged, the *Gm* genome has undergone a whole genome duplication event, estimated to have occurred 10 to 14 Mya (Shoemaker et al., 2006). As expected, we identified two homeologous regions syntenic with the *Pv* B4 410-kb contig in the 1.01 soybean genome database (http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01/). The first one (referred to as H1, for homeologous 1 *Pv* B4 syntenic region) is divided into chromosome 13 (coordinates 27,096,894–27,149,916 bp) and chromosome 5 (coordinates 8,372,893–8,464,947 bp), and the second one (referred to as H2, for homeologous 2 *Pv* B4 syntenic region) corresponds to chromosome 19 (coordinates 704,347–861,494 bp; Supplemental Fig. S3). Eighteen out of the 25 low-copy genes in the *Pv* B4 410-kb contig have homologs in the *Gm* H1 *Pv* B4 syntenic region,

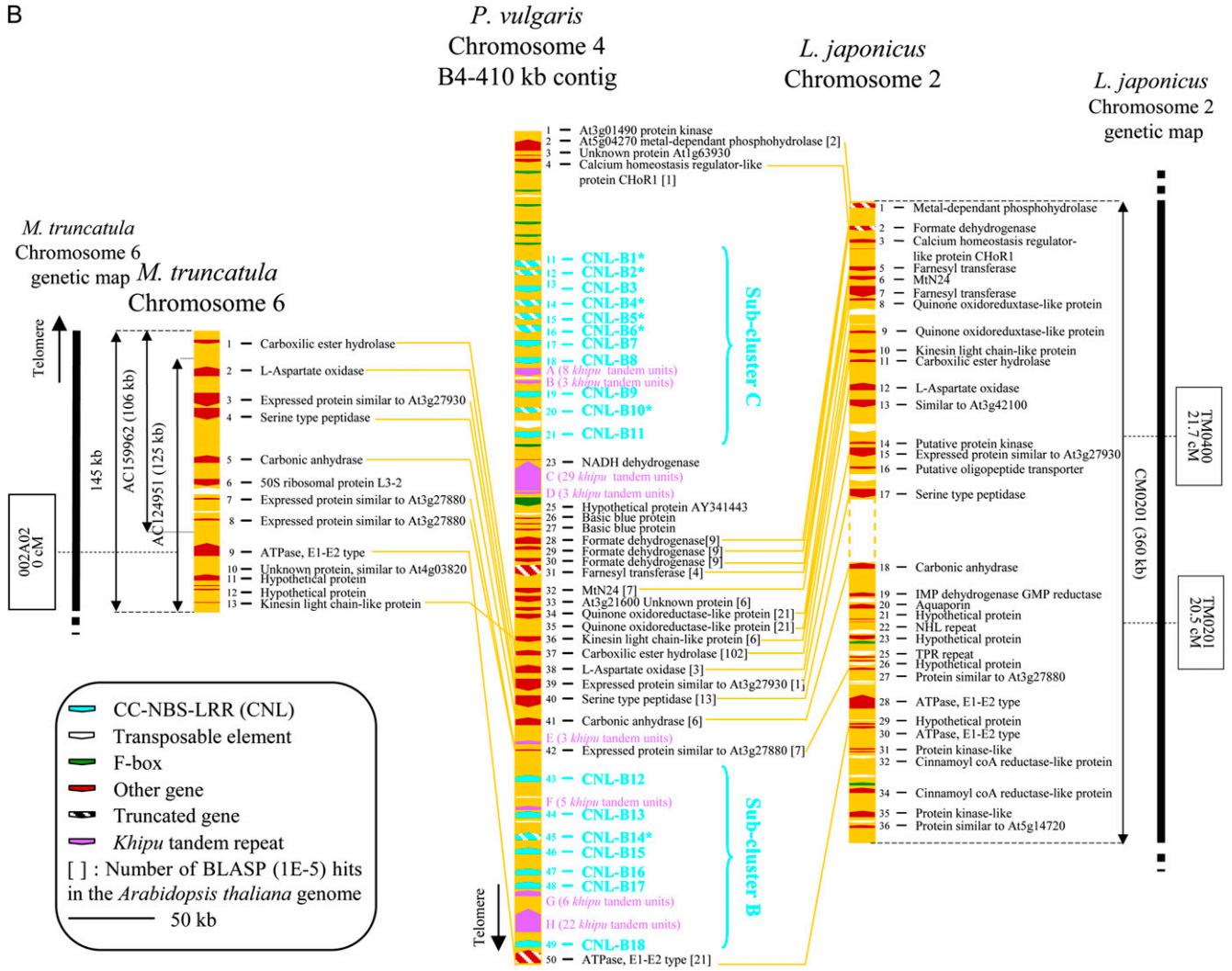

Figure 1. Schematic representation of 650 kb of the *Pv* BAT93 B4 *R* gene locus sequence. A, Sequence annotation of the 48-B10 and FZ-E9 BAT93 BAC clones and of the 410-kb contig (four overlapping BAC clones: 75-H11, 254-G15, FY-N24, and 94-L19). Two gaps are indicated by dotted lines. The orientation of predicted genes is indicated by arrows. Putative gene function is listed on the right. Twenty-six CNL (turquoise arrows) are regrouped in four subclusters named A to D. The *khipu* tandem repeats (A–L) are in purple. Genetic positions are on the right (GP 1–GP 4), as described in the corresponding physical map presented by David et al. (2008). Genetic distances between each GP are indicated on the right in Kosambi cM. BAC clones FZ-E9 and 48-B10 have been mapped at GP 4 and 1, respectively. Orientations of these two BAC clones are unknown, so arbitrary orientations were chosen for this schematic representation. B, Sequence comparison between the *Pv* BAT93 B4 410-kb contig (center) and syntenic regions in *Mt* chromosome 6 (left) and *Lj* chromosome 2 (right). Yellow lines indicate significant homology matches between predicted genes. BAC clone position along the sequence is to scale. Genetic maps of these syntenic regions are presented on the left of the *Mt* syntenic region and on the right of the *Lj* syntenic region. Molecular markers and their genetic distances (cM) in *Mt* and *Lj* genetic maps are presented in boxes on the left and right, respectively.

and 17 have homologs in the *Gm* H2 *Pv* B4 syntenic region, mostly conserved in order and orientation. Contrary to what we observed in *Mt* and *Lj*, both the *Gm* H1 and H2 *Pv* B4 syntenic regions contain a single truncated CNL, with approximately 80% nucleotide identity to the *Pv* B4-CNL (Supplemental Fig. S4). These two *Gm* CNL are collinear to the *Pv* B4-CNL from subcluster B, whereas CNL from subcluster C are completely absent in both *Gm* H1 and H2 *Pv* B4 syntenic regions (Supplemental Fig. S3).

In conclusion, clear syntenic regions of the *Pv* B4 cluster were identified in the *Mt*, *Lj*, and *Gm* genomes. The average amino acid similarity between *Pv* B4 genes and corresponding *Lj*, *Mt*, and *Gm* orthologous genes are 70.4%, 70.8%, and 75%, respectively (Supplemental Table S6). In sharp contrast to the collinearity observed for non-CNL-encoding genes, *Pv* B4-CNL sequences were completely absent in *Mt* and *Lj* syntenic regions, and only a truncated CNL is present in the two subcluster B syntenic regions of *Gm*. All these

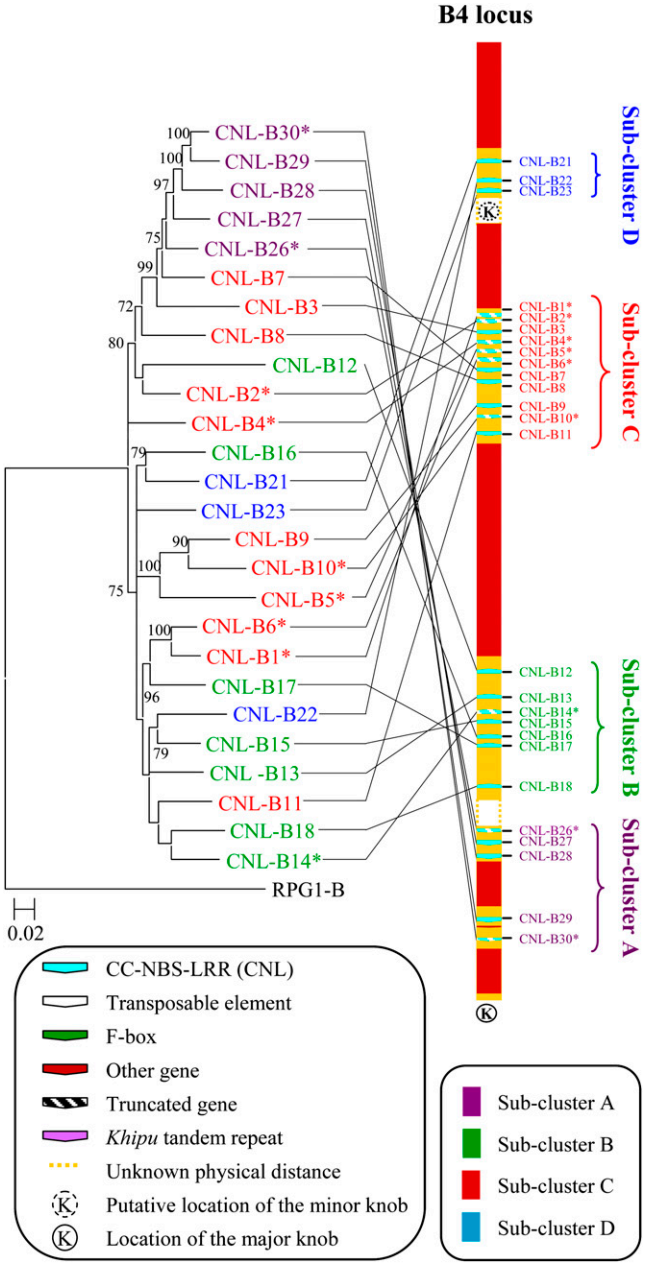

Figure 2. Bayesian phylogenetic tree of BAT93 *Pv* B4-CNL sequences. This tree was constructed using the complete NBS domains (from P-loop to MHDV) of the 26 BAT93 B4-CNL nucleic acid sequences presented in Figure 1. Numbers on the branches represent posterior probabilities (only >70% are indicated) after running a Markov Chain Monte Carlo search for 10,000,000 generations. Physical positions of the B4-CNL along the *B4* locus are indicated on the right. The soybean *Rpg1-b* sequence was used as an outgroup.

data suggest that these CNL either were inserted in the ancestor of the B4 locus after *Mt*, *Lj*, and *Pv* diverged but before the divergence of *Pv* and *Gm* ("gain" hypothesis) or, alternatively, were lost from *Mt* and *Lj* after the divergence of *Pv*, *Gm*, *Mt*, and *Lj* ("loss" hypothesis).

Pv B4-CNL Derived from CNL of the *Pv* Co-2 Gene Cluster

In order to test the gain and loss hypotheses, we searched if *Pv* B4-CNL homologs exist elsewhere in the *Mt*, *Lj*, and *Gm* genomes. We performed TBLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi) analysis using *Pv* B4-CNL as a query against the three sequenced legume species genomes. This analysis led to the identification of two *Mt*, two *Lj*, and 29 *Gm* full-length CNL sequences (0 BLAST e-value), presenting a minimum of 60% amino acid similarity with *Pv* B4-CNL sequences. In soybean, the 29 identified full-length CNL are all members of the *Rpg1-b* resistance cluster: four are alleles of the *Rpg1-b* disease resistance gene (Ashfield et al., 2004), three are members of the *Rpg1-b* locus (genotype PI96983; Hayes et al., 2004), and 19 and three are members of the H1 and H2 regions of the *Rpg1-b* cluster (genotypes Williams 82 and PI96983; Innes et al., 2008), respectively. In *Lj*, the two identified CNL sequences (*Lj3fg26933* and *Lj3fg26877*) are separated by 18.2 kb on chromosome *Lj3* (BAC clone AP009688). Finally, in the *Mt* genome, the two CNL sequences AC153125_16.5 (referred to as *Mt8g1223*) and AC153125_27.5 (referred to as *Mt8g1225*) are separated by 11.3 kb on chromosome *Mt8* (BAC clone AC153125; Supplemental Table S5). Previous comprehensive analysis including all identified *Mt* CNL sequences has revealed that *Mt8g1223* and *Mt8g1225* belong to a small clade (referred to as CNL-1) that diverged after the legume/poplar split comprising only five *Mt* CNL members: locally duplicated genes *Mt8g1223* and *Mt8g1225* on chromosome 8 and locally duplicated genes *Mt2g1873* (AC148761_18.3), *Mt2g1878* (AC148761_27.3), and *Mt2g1862* (AC148761_2.3) on chromosome 2 (Ameline-Torregrosa et al., 2008; Innes et al., 2008). Separation of these chromosomal regions may date to a whole genome duplication that occurred early in legume evolution (Shoemaker et al., 2006). Furthermore, the regions including the two identified *Mt8* sequences and the soybean *Rpg1* cluster are known to be syntenic (Innes et al., 2008). Consequently, putative orthologs of the B4-CNL are present in the *Mt*, *Lj*, and *Gm* genomes but in nonsyntenic regions of the B4 *R* gene cluster.

To investigate the relationship between *Pv* B4-CNL and these 33 most similar sequences in *Mt*, *Lj*, and *Gm*, we conducted phylogenetic analyses including those legume CNL sequences and all known *Pv* B4-CNL sequences from BAT93 and from another genotype, JaloEEP558 (Ferrier Cana et al., 2003; Geffroy et al., 2009). We also added two representative *Mt* CNL sequences per *Mt* clade CNL-2 to -5 as well as two Arabidopsis and six poplar CNL sequences as nearest outgroups to the CNL-1 clade described by Ameline-Torregrosa et al. (2008). Furthermore, 21 additional *Gm* Rpg1-CNL sequences and 20 *Pv Rpg1-b* orthologs described by Innes et al. (2008) were also included in this analysis (Fig. 3). In order to know the genomic location of these *Rpg1* orthologs (Innes et al., 2008) in

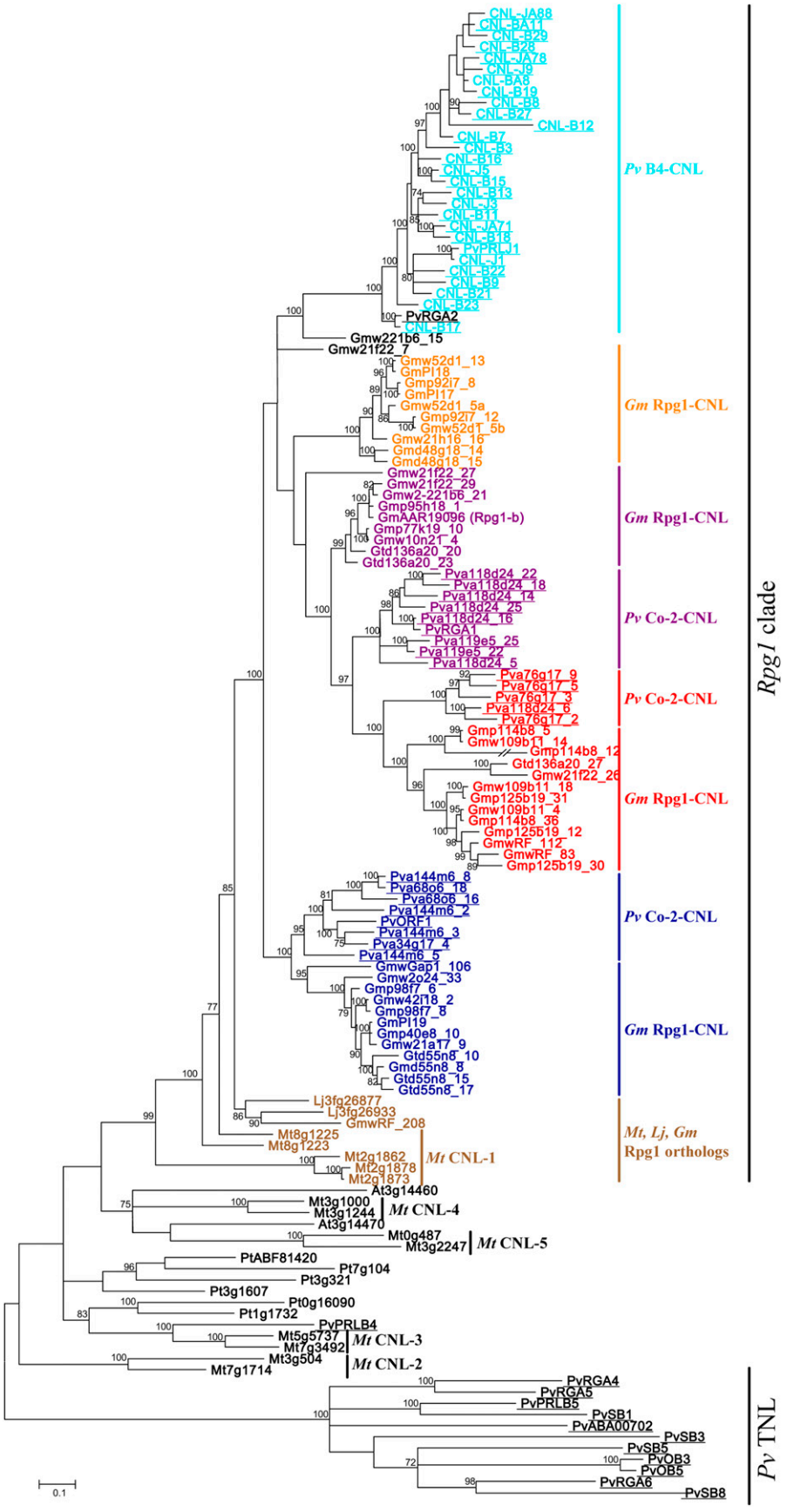

Figure 3. Phylogenetic analysis of NL genes. This Bayesian tree was constructed using just the NBS domain (from the P-loop to the GLPL motif) of 133 amino acid sequences. Numbers at nodes indicate posterior probabilities (only >70% are indicated). Gene name abbreviations are as follows: Pv, *Phaseolus vulgaris*; Gmw, *Glycine max* Williams 82; Gmp, *Glycine max* line PI96983; Gtd, *Glycine tomentella* diploid accession G1403; Mt, *Medicago truncatula* var Jemalong; Lj, *Lotus japonicus* accession Miyakojima MG-20; At, *Arabidopsis thaliana*; Pt, *Populus trichocarpa*. *Rpg1* orthologs, described by Innes et al. (2008), are regrouped into one clade; *Pv* B4-CNL (in turquoise) belong to the *Rpg1* ortholog clade. *Rpg1* orthologs are colored as described by Innes et al. (2008). The five *Mt* CNL clades (*Mt* CNL-1–*Mt* CNL-5) and eight poplar and Arabidopsis NL sequences are as described by Ameline-Torregrosa et al. (2008). All *Pv* NL sequences available in the nr database have been included in this analysis and are underlined on the tree. The tree was rooted using *Pv* TNL sequences. Sequence characteristics are reported in Supplemental Table S5.

the *Pv* genome, we have developed a genetic mapping strategy. It revealed that these *Pv Rpg1* orthologs mapped at the end of LG B11 in the vicinity of a previously identified resistance locus, the Co-2 locus (Geffroy et al., 1998; Creusot et al., 1999). Consequently, they will be referred to as *Pv* Co-2-CNL. Finally, TBLASTX (http://blast.ncbi.nlm.nih.gov/Blast.cgi) searches have been carried out to include in our phylogenetic analysis all *Pv* NL sequences available in GenBank. This led to the identification of 16 *Pv* sequences in addition to *Pv* Co-2-CNL and *Pv* B4-CNL.

A well-supported clade (bootstrap value of 99) is composed, as expected, of *Lj3fg26933*, *Lj3fg26877*, *Mt8g1223*, *Mt8g1225*, *Mt2g1873*, *Mt2g1878*, and *Mt2g1862* and orthologous soybean Rpg1-CNL and *Pv* Co-2-CNL sequences (Fig. 3). Moreover, *Pv* B4-CNL sequences form a distinct subclade in this major clade, referred to as "*Rpg1* clade." Together, these data indicate that *Pv* Co-2-CNL and *Pv* B4-CNL have the same orthologs in the *Mt*, *Lj*, and *Gm* genomes, even though they are located at two different genomic locations in the bean genome: at the ends of LG B11 and LG B4. Considering that *Pv* Co-2-CNL, *Mt8g1223*, and *Mt8g1225* and soybean Rpg1-CNL are located in syntenic regions between *Pv*, *Mt*, and *Gm*, and considering that this is not the case for *Pv* B4-CNL sequences, we deduce that *Pv* B4-CNL was derived from the *Pv* Co-2 locus (gain hypothesis). In the loss hypothesis, B4-CNL sequences (turquoise clade) would have been present prior to the *Mt*, *Lj*, *Pv*, and *Gm* divergence and then would have been deleted in a *Mt* and *Lj* ancestor. Given the topology of the tree presented in Figure 3, this would imply that the turquoise clade and the clade comprising the orange, red, and purple sequences would have diverged prior to the *Mt*, *Lj*, *Pv*, and *Gm* divergence and also that the blue clade diverged from these two clades prior to the *Mt*, *Lj*, *Pv*, and *Gm* divergence. Given that no *Mt* or *Lj* CNL sequences fall into any of these three clades, the loss hypothesis could only be explained by at least three distinct events of CNL extinctions in a *Mt* and *Lj* ancestor. Thus, the gain hypothesis, which implies only one event of CNL birth in a *Pv* and *Gm* ancestor, is more parsimonious than the loss hypothesis. Consequently, although the tree topology is complicated, it supports the gain hypothesis.

Together, these results support the hypothesis that the *Pv* B4-CNL were derived from CNL of the *Pv* Co-2 cluster through an ectopic recombination event between nonhomologous chromosomes that occurred after the divergence from a common ancestor of *Lj*, *Mt*, and *Pv* but before *Pv* and *Gm* diverged.

Identification of a *Phaseolus*-Specific Satellite Sequence Located in a Subtelomeric Region of Bean Chromosomes, the *khipu* Tandem Repeat

In addition to CNL sequences, annotation of the 645 kb of the *Pv* B4 locus revealed the presence of a satellite DNA in the 48-B10 BAC clone and in the 410-kb contig but absent in the FZ-E9 BAC clone. Indeed, we have identified 92 very well-conserved repeat units (Supplemental Fig. S5), arranged in tandem array in a head-to-tail orientation, and referred to as *khipu*. The 92 *khipu* units vary in length from 311 to 564 bp, and the sequence identity among the 92 *khipu* units ranged from 40.2% to 99.8%, with an average of 80.1% (Supplemental Fig. S6). A consensus nucleotide sequence of 528 bp with a GC content of 38.2% was defined from multiple alignment of the 92 *khipu* units. Twelve *khipu* tandem repeats (A–L) have been identified. The number of *khipu* units in a repeat varies from 1 (repeat I) to 29 (in tandem repeat C; Supplemental Fig. S1). Interestingly, *khipu* tandem repeats are not randomly distributed on the 645 kb but are tightly interspaced between CNL sequences: tandem repeats A to D are localized in the *Pv* B4-CNL subcluster C, while tandem repeats E to H and I to L are located in the *Pv* B4-CNL subclusters B and A, respectively. No *khipu* tandem repeat was identified in the *Pv* B4-CNL subcluster D. Within a tandem repeat, all *khipu* units have the same orientation, while within a CNL subcluster, all *khipu* tandem repeats do not have the same orientation. For example, in the *Pv* B4-CNL subcluster C, *khipu* tandem repeats A, B, and D are oriented in the same direction while *khipu* tandem repeat C is oriented in the opposite direction (Supplemental Fig. S1). The same observation was made for *khipu* tandem repeats belonging to *Pv* B4 CNL subcluster A.

In order to infer the chromosomal location of the *khipu* tandem repeats in the BAT93 common bean genotype, FISH analysis has been performed. First, the hybridization of the BAC clone ER-N1 containing *Pv* B4-CNL sequences from subcluster C (David et al., 2008) on mitotic chromosome (Fig. 4A) gives a clear signal only at the end of the short arm of chromosome 4, indicating that this BAC clone ER-N1 probe is specific for chromosome 4 (this ER-N1 BAC clone probe is referred to as *Pv* B4-CNL probe). A second hybridization experiment using *khipu* as probe has been carried out, and the *khipu* probe hybridized to most of the chromosome ends, including both ends of chromosome 4 (Fig. 4B). Although the strongest signals were detected on the short arm of chromosome 4, *khipu* repeats are not restricted to the B4 *R* gene cluster. In order to determine the pattern of *khipu* distribution at greater resolution, FISH was also performed on meiotic pachytene chromosomes because they are less condensed than somatic chromosomes. Terminal knobs of different sizes were visible at most chromosome ends (Fig. 4C), while *khipu* tandem repeats were present on 17 chromosome ends, mostly corresponding to cytologically visible terminal or subterminal knobs (Fig. 4D). Furthermore, signal intensity correlated with knob size at most chromosome ends (white arrowheads in Fig. 4D), suggesting that *khipu* is an important component of the considered knobs. No detectable hybridization of *khipu* was found for five chromosome ends (Fig. 4, C and D), despite the

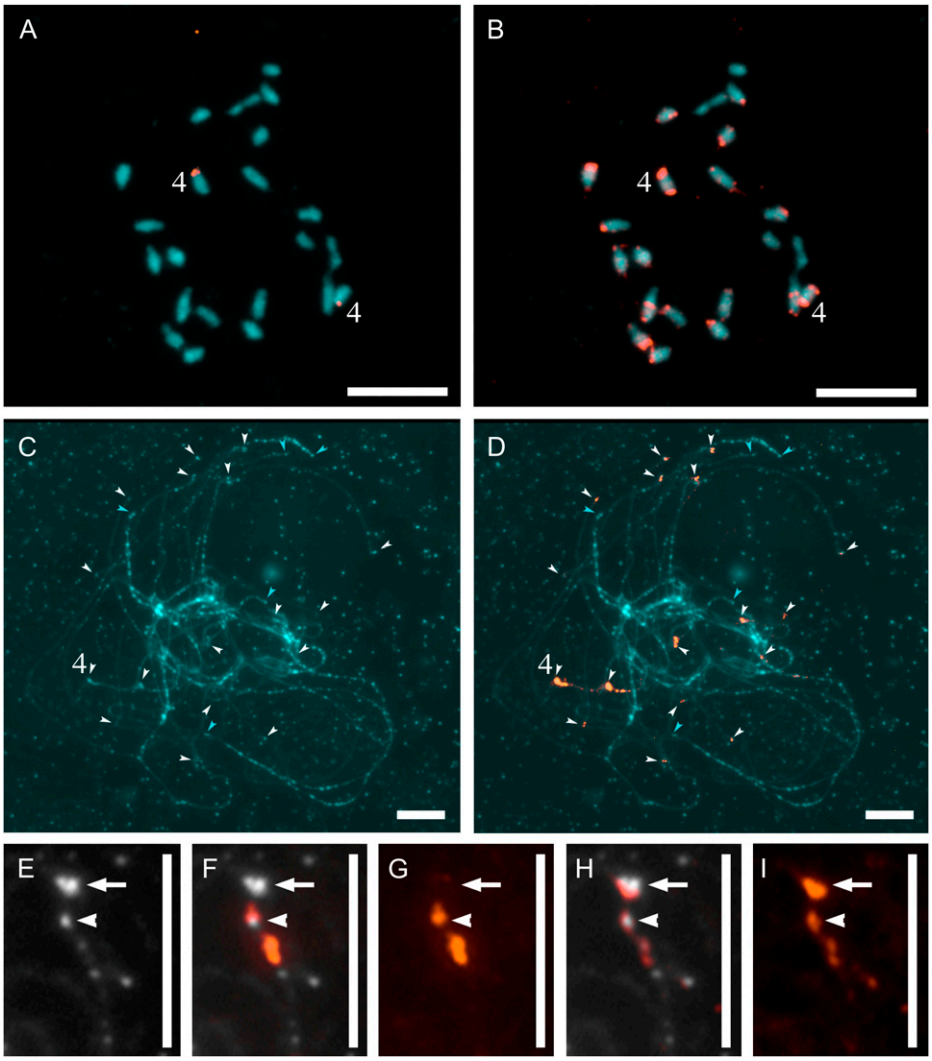

Figure 4. FISH to *Pv* BAT93 chromosomes. A and B, FISH to mitotic chromosomes. Signals are in red, and chromosomes are in blue. Chromosome 4 is indicated. A, ER-N1 BAC clone (David et al., 2008) was used as a probe. ER-N1 BAC, containing *Pv* B4-CNL sequences from subcluster C of the *Pv* B4 410-kb contig, is specific for the end of the short arm of chromosome 4. B, *Khipu* was used as a probe. A repetitive distribution pattern is observed, including strong signals on terminals of both chromosome 4 arms. C to I, FISH to pachytene chromosomes. C and D, All chromosome ends are tentatively indicated by arrowheads. *Khipu* was used as a probe in D. Note that although signal intensity was correlated with knob size in most chromosomes ends (white arrowheads), not all terminal knobs are labeled with the *khipu* probe (blue arrowheads). Two strong hybridization signals are detected on the end of the short arm of chromosome 4, which is tentatively indicated. E to I, Focus on the short arm of chromosome 4. The major knob and the minor knob are indicated with a white arrow and a white arrowhead, respectively. F and G, ER-N1 BAC clone was used as a probe and is shown in red superimposed on the DAPI-stained chromosome (F) or not (G). Note signals on both sides of the minor knob. H and I, *Khipu* was used as a probe and is shown in red superimposed on the DAPI-stained chromosome (H) or not (I). Note signals on minor and major knobs as well as on adjacent euchromatin. Bars = 10 μm.

presence of a knob, suggesting that *khipu* is not the only repeated sequence present in bean knobs. Furthermore, *khipu* signals were not restricted to the terminal knobs, especially on terminal knobs showing strong labeling, but were spread over the euchromatic regions adjacent to the knobs (Fig. 4D). As defined by Geffroy et al. (2009), two knobs referred to as the minor knob (white arrowhead) and major knob (white arrow) are visible at the B4 *R* gene cluster. Hybridization with the *Pv* B4-CNL probe revealed clear signals on both sides of the minor knob located at the end of chromosome 4 (Fig. 4, E–G), confirming previous reports that *Pv* B4-CNL subclusters are localized on both sides of this minor knob (Geffroy et al., 2009). The *khipu* tandem repeats revealed a strong association with the major and minor knobs, demonstrating that *khipu* is an important component of these two knobs (Fig. 4, D H, and I). Finally, colocalization of the *khipu* and *Pv*

B4-CNL probe signals was observed, confirming our sequence data indicating that *khipu* tandem repeats are found within the *Pv* B4-CNL subclusters (Fig. 4, F–I).

Searches of sequence databases with BLAST did not reveal any significant homology of the consensus *khipu* sequence to any previously described plant satellite DNA. Phylogenetic origin of *khipu* sequence was studied by looking at the distribution of the sequence in related species. The *khipu* tandem repeats were used as a probe in Southern hybridization experiments of digested genomic DNA from various *Phaseolus* species (*P. glabellus*, *P. microcarpus*, *P. oligospermus*, *P. pedicellatus*, *P. filiformis*, *P. parvifolius*, *P. vulgaris*, *P. dumosus*, *P. leptostachyus*, and *P. lunatus*) as well as from six other legumes (*Mt*, *Centrosema macrocarpum*, *Teramnus uncinatus*, *Vigna luteola*, *V. vexillata*, and *Macroptilium atropurpureum*). The *khipu* tandem repeats were detected in all *Phaseolus* species tested, while no signal was detected with any of the six other tested legume species (Supplemental Fig. S7). Consequently, high-copy sequences fairly homologous to *khipu* are only present in species belonging to the *Phaseolus* genus. The copy number of *khipu* tandem repeats varies from one *Phaseolus* species to another, as testified by the differential intensity of hybridization signals: *P. dumosus*, *P. vulgaris*, and *P. parvifolius* (all from the Vulgaris group according to Delgado-Salinas et al., 2006) appeared to have the highest copy numbers, while *P. oligospermus*, *P. pedicellatus*, *P. leptostachyus*, and *P. lunatus* seemed to have the lowest copy numbers. In order to know if there was a strict correlation between the presence of *khipu* tandem repeat and *Pv* B4-CNL in the legume family, Southern analysis has been carried out with the NBS *PRLJ1* probe, specific to the *Pv* B4 *R* cluster. Weak signals for *V. luteola* and *V. vexillata* and stronger signal for *M. atropurpureum* have been detected. Strongest signals have been detected for the other *Phaseolus* species except for *P. filiformis* (Supplemental Fig. S7). In *V. luteola*, *V. vexillata*, and *M. atropurpureum*, a *PRLJ1* signal was identified independently of a *khipu* signal, suggesting that *Pv* B4-CNL appeared before the *khipu* tandem repeat appeared in legumes. Because of the presence of the *khipu* tandem repeats only in *Phaseolus* sensu stricto species (and the absence in the Phaseoleae-Phaseolinae [*Macroptilium* and *Vigna*], the Phaseoleae-Clitoriinae [*Centrosema*], the Phaseoleae-Glycininae [*Teramnus*], and the Trifolieae [*Medicago*]), our results would thus be another confirmation that *Phaseolus* is monophyletic and that its genesis is not made of the addition of morphologically similar and convergent neotropical legumes (Chacon et al., 2007).

DISCUSSION

By exploiting the resources generated by the *Mt*, *Lj*, and *Gm* genome projects in a targeted comparative study, centered on approximately 645 kb spanning the *Pv* B4 *R* gene cluster, and by combining genomic, phylogenetic, and cytogenetic approaches, we gained insights into both *R* gene evolution and genome plasticity. In current models of the evolution of plant disease *R* genes, the source of genetic variability is mainly based on mechanisms such as unequal crossing over, gene conversion, and point mutations. In contrast, our results point out the importance of an additional mechanism for *R* gene evolution: ectopic recombination between nonhomologous chromosomes. In particular, we propose that subtelomeric dynamics might fuel rapid changes in subtelomeric gene repertoires, in agreement with the need of plant *R* genes to evolve rapidly to cope with an ever-changing array of pathogens.

Our in-depth analysis of the region orthologous to the end of *Pv* chromosome 4 in three legume genomes (*Lj*, *Gm*, and *Mt*; Fig. 1B; Supplemental Fig. S3) revealed clear syntenic regions as well as CNL sequence movement. Considering non-CNL-encoding genes, extensive gene collinearity was observed across the legume family, confirming previous syntenic relationships such as between parts of *Lj2* and *Mt6* (Zhu et al., 2005; Cannon et al., 2006). Considering that the most recent common ancestor between *Pv, Lj, Gm*, and *Mt* is dated 54 Mya (Lavin et al., 2005; Cronk et al., 2006), the level of microsynteny observed between the four species is remarkable. Interestingly, this collinearity was strictly restricted to non-CNL-encoding genes, suggesting the existence of CNL gene movement. Since legumes are the only plant family with three genomes sequenced or nearly completely sequenced, we had the unusual opportunity to infer the origin of the *Pv* B4-CNL using a combination of comparative genomic and phylogenetic approaches. These analyses suggested that the *Pv* B4-CNL from LG B4 was derived from CNL from another *R* cluster, the Co-2 cluster, located at the end of LG B11 (Geffroy et al., 1998), through an ectopic recombination event. This phylogenetic proximity between *Pv* B4-CNL and *Pv* Co-2-CNL is supported by an independent experiment based on NL mapping in the *Pv* genome, where a CNL probe (referred to as RGA2) was mapped at both *Pv* B4 and *Pv* Co-2 *R* gene loci, confirming that *Pv* B4-CNL and *Pv* Co-2-CNL (presenting on average 65% of nucleotide identity in the NBS domain) are the two most similar CNL families in the *Pv* genome (Lopez et al., 2003). Integration of the soybean genome data enables us not only to more precisely date this event but also to infer that a single CNL moved between these two *R* gene loci. Indeed, a single truncated CNL sequence is present in both soybean H1 and H2 B4 subcluster B syntenic regions (Supplemental Fig. S3). These data strongly suggest that only one CNL sequence was implicated in the gene movement from the ancestral Co-2 cluster to the ancestral B4 cluster and that subcluster B corresponds to the ancestral part of the B4-CNL cluster. Since a CNL sequence has been found in both soybean H1 and H2 B4 cluster syntenic regions but is absent in the corresponding *Mt* and *Lj* regions, we conclude that this ectopic recombination event occurred after *Mt*, *Lj*, and *Pv* diverged (54 Mya)

but before *Pv* and *Gm* diverged (19 Mya; Lavin et al., 2005; Cronk et al., 2006). Our results, centered on an *R* gene cluster in legumes, confirm recent large-scale sequence analysis. Indeed, comparative genomic analyses taking into account complete genome sequences have revealed massive gene movement between chromosomal positions (Ammiraju et al., 2008; Freeling et al., 2008; Lu et al., 2009). In particular, NL sequences were shown to be particularly prone to move. Freeling et al. (2008) report that 91% of NL sequences have moved to a new chromosomal position since the origin of the order Brassicales. It is unclear what portion of this "movement" is due to transposition versus segmental duplications and erasure of synteny in rapidly evolving NL clusters (Baumgarten et al., 2003), but the net effect is pronounced: the majority of NL genes lack syntenic context. Analysis of synteny of *Arachis* with *Lotus* and *Medicago* confirms this tendency (Bertioli et al., 2009). Furthermore, these studies do not illuminate the potential molecular mechanism underlying these gene movements.

We have presented additional evidence for the existence of ectopic recombinations between nonhomologous chromosomes in common bean. This has been possible thanks to the identification of a new 528-bp satellite DNA referred to as *khipu*, which is mostly associated with bean chromosome ends, in subtelomeric regions (Fig. 4). Indeed, the occurrence of the *khipu* tandem repeat in most chromosome ends, while *khipu* is specific to the genus *Phaseolus* (Supplemental Fig. S7), provides undisputed evidence of frequent sequence exchange between *Pv* nonhomologous chromosome ends. The B4 *R* gene cluster is located in a subtelomeric region of the short arm of chromosome 4. The subtelomeric location of the Co-2 locus is known indirectly, since the Co-2 locus mapped at the end of LG B11, in a distal position relative to a conserved legume marker (Leg220; Hougaard et al., 2008), which is known to be located terminally in chromosome 11 (A. Fonsêca and A. Pedrosa-Harand, unpublished data). Thus, one model to explain the origin of B4-CNL is that the sequence exchange between *Pv* nonhomologous chromosomes could have occurred in the context of frequent sequence exchanges between nonhomologous subtelomeric chromosome ends. This model is supported by the results obtained in other organisms such as human, yeast, and trypanosome, where subtelomeres are known to be hot spots of interchromosomal recombination (Mefford and Trask, 2002; Linardopoulou et al., 2005). In human, bioinformatics analyses highlighted the role of the nonhomologous end-joining DNA repair process in the generation of these patchwork structures in subtelomeric regions (Linardopoulou et al., 2005). Such plastic regions of the genome are supposed to be evolutionarily dynamic. Interestingly, multigenic families located in subtelomeric regions have been associated with adaptive processes in yeast, human, and *Plasmodium falciparum* (Charron et al., 1989; Turakainen et al., 1993; Su et al., 1995; Mefford and Trask, 2002; Linardopoulou et al., 2005). Consequently, our study not only confirms the existence of CNL movement in plant genomes but also provides a possible explanation of the molecular basis of this gene movement. Analyses of subtelomeric regions in plants are scarce (Kuo et al., 2006), and to our knowledge, this peculiar feature of subtelomeres has not yet been reported in plants. In a recent study, indirect evidence of gene movement in rice subtelomeres comes from the comparative genomics analyses of a subtelomeric region between three *Oryza* species (Fan et al., 2008). In plants, further indirect evidence of ectopic recombination between nonhomologous chromosomes in subtelomeric regions has been provided by FISH analysis in rye (*Secale cereale*). Indeed, heterochromatin connections between nonhomologous chromosomes were identified in subtelomeric regions within a satellite DNA sequence (Gonzalez-Garcia et al., 2006). The existence of ectopic recombination in subtelomeric regions is also supported by results with 45S rDNA variation in common bean (Pedrosa-Harand et al., 2006). Interestingly, in the *Pv* genome, most of the large NL clusters have been located at the end of linkage groups. For example, *Co-x*, *I*, *Co-4*, Co-2, and B4 resistance loci are located at the ends of LG B1, B2, B8, B11, and B4, respectively (Melotto et al., 2004; Vallejos et al., 2006; Geffroy et al., 2009), and a subtelomeric location has been demonstrated or inferred for *Co-4*, B4, and Co-2 loci. Thus, regarding the nonrandom repartition of NL sequences in *Pv*, it is tempting to speculate that NL gene movement in subtelomeric regions has played an important role in the evolution of NL clusters in *Pv*. In the case of plant *R* gene clusters, since plants have to maintain a genetic diversity necessary to recognize a plethora of effectors, ectopic recombination events leading to the dispersal of duplicated copies to physically distant sites could provide a mechanism by which genetic novelty could escape erasure by concerted evolution (McDowell and Simon, 2006). However, subtelomere plasticity is not the only mechanism involved in NL movement in plant genomes, as exemplified by the Arabidopsis genome, where many NL clusters are not subtelomeric (Meyers et al., 2003).

In *Gm*, a B4-CNL homolog has been identified, although no *khipu* homolog has been identified. Furthermore, hybridization analyses on various legume species showed no perfect correlation between signal intensity detected with *PRLJ1* (B4-CNL) and *khipu* probes (Supplemental Fig. S7). These results could suggest that the sequence movement involving *khipu* and the B4-CNL sequences corresponds to two independent events. Since three species present a *PRLJ1* signal but no *khipu* signal (in agreement with what is observed in *Gm*), the CNL movement could have occurred first and the *khipu* movement could have occurred at a second time. However, an essential feature of tandem arrays of satellite DNA is that they are considered fast-evolving parts of the eukaryotic genome (Waye and Willard, 1989; Vershinin et al., 1996; Alexandrov et al., 2001). Thus, within a species

such as *V. luteola*, which exhibits no *khipu* hybridization signal (Supplemental Fig. S7), a *khipu*-related satellite may be present but too divergent to be detected by our *khipu* probe.

Subsequent to this ectopic recombination event, phylogenetic and genomic data indicate that tandem and ectopic duplications (Fig. 2) have occurred within the B4-CNL locus and led to rapid amplification of the *Pv* B4-CNL sequences during the *Phaseolus* genus emergence (Fig. 2; Supplemental Fig. S7). This amplification is impressive, starting from one CNL sequence in subcluster B less than 19 Mya, and leading to at least 29 CNL sequences organized into four subclusters today (Ferrier Cana et al., 2003; Geffroy et al., 2009; this study). Consequently, the *Pv* BAT93 B4 *R* gene cluster is one of the largest clusters identified so far in plants and is comparable to the lettuce (*Lactuca sativa*) *Dm3* locus, which is composed of 32 members within 3.5 Mb (Meyers et al., 1998; Kuang et al., 2004). Important variations in the number of gene copies (from one to 52 copies) have been estimated among haplotypes of the maize (*Zea mays*) *Rp1* cluster (Smith et al., 2004), in agreement with the fact that the *R* gene cluster can expand and possibly shrink rapidly within short time periods. Our results not only characterize the origin of the *Pv* B4-CNL cluster but also point out the rapid birth of a novel, large, and complex subtelomeric CNL cluster. We use the term "nomadism" to describe this property of NL sequences to migrate from an ancestral genomic location to another one and then to proliferate locally. Numerous specific *R* genes and *R* quantitative trait loci effective against a broad range of pathogens map at the B4 locus (Geffroy et al., 1999, 2000; Lopez et al., 2003; Miklas et al., 2006). Moreover, five B4-CNL were found to be expressed (Ferrier Cana et al., 2003). Altogether, our data suggest that NL nomadism is capable of generating novel resistance genes in a short evolutionary time span and should be considered as an important evolutionary mechanism for the evolution of disease resistance genes.

Based on our comparative genomics, phylogenetic, and cytogenetic data, as well as previous synteny analyses between the *Pv*, *Lj*, *Gm*, and *Mt* genomes (Choi et al., 2004; Zhu et al., 2005; Cannon et al., 2006; Hougaard et al., 2008; Tsubokura et al., 2008), we propose a model of evolution of the B4 and Co-2 loci from the common ancestor of *Mt*, *Lj*, *Gm*, and *Pv* to the current state in agreement with the gain hypothesis (Supplemental Fig. S8). We propose that ancestral Co-2 and B4 loci (referred to as loci aCo-2 and aB4, respectively) were located on two nonhomologous chromosomes in the progenitor, given that chromosome macrosynteny is observed between *Pv11* and *Lj3* as well as between *Pv4* and *Lj2* along their whole length (Hougaard et al., 2008). We also propose a subtelomeric location of aCo-2 and aB4, because of the subtelomeric location of the *Pv* B4 locus and its corresponding syntenic regions on *Mt6* and *Gm19* and because of the inferred subtelomeric location of the Co-2 locus on *Pv11*. We present three scenarios of evolution of the aCo-2 and aB4 loci, including the CNL ectopic recombination event from the aCo-2 locus to the aB4 locus (between 54 and 19 Mya). Each scenario highlights the role of subtelomere plasticity in the evolution of these two loci. We present for each scenario the simplest series of events to explain the current state of Co-2 and B4 loci. The most likely scenario corresponds to scenario C, considering that it implies both (1) the fewest events in *Pv* and (2) the most likely series of events in *Gm* to explain the contiguous location of orthologous Co-2 and B4 loci on chromosome *Gm13*. Even if scenario C is not the most parsimonious scenario in *Gm*, it is supported by recent discoveries showing that *Gm* LGs consist of mosaics of relatively small segments syntenic with *Lj* (Tsubokura et al., 2008), indicating that massive rearrangement has occurred in *Gm* after the whole genome duplication event estimated to have occurred 10 to 14 Mya (Shoemaker et al., 2006).

Unlike species with large genomes, heterochromatin is usually restricted to pericentromeric regions in small-genome plant species. For example, in the compact Arabidopsis genome (125 Mb; Arabidopsis Genome Initiative, 2000) only two knobs have been reported (Fransz et al., 2000), while in the large maize genome (2,671 Mb; Bennett and Smith, 1991), numerous knobs have been reported (Peacock et al., 1981; Ananiev et al., 1998). The relatively small genome of *Pv* does not seem to follow this rule. Indeed, heterochromatic knobs have been detected in most *Pv* chromosome termini, and we have demonstrated that *khipu* tandem repeats are components of most of them (Fig. 4). This abundance of terminal knobs in *Pv* is in sharp contrast to results from other legume species such as *Lj* and *Mt*, where most of the heterochromatin is localized at pericentromeric regions and no terminal heterochromatic blocks have been reported, except for the 45S rDNA cluster on *Lj2* (Kulikova et al., 2001; Pedrosa et al., 2002). Thus, the complexity of bean subtelomeres does not seem to be obviously related to its genome size, because at 588 Mb (Bennett and Leitch, 1995), *Pv* is not significantly larger than *Lj* (472 Mb; Sato et al., 2008) or *Mt* (500 Mb; Kulikova et al., 2001). Our results confirm, therefore, a peculiarity of the *Pv* genome and constitute additional evidence that genome size alone cannot predict genome structure in plants. Previously reported peculiar features of the *Pv* genome are its richness in repetitive sequences (Schlueter et al., 2008) compared with other legume species presenting a larger genome size, such as *Trifolium repens* (956 Mb; Bennett and Smith, 1991) and soybean (1,103 Mb; Bennett and Leitch, 1997), and its less compartmentalized nature (Pedrosa-Harand et al., 2009). Indeed, a large proportion of BAC clones selected with single-copy molecular markers showed pericentromeric, subtelomeric, or dispersed repetitive patterns, demonstrating that the common bean genome presents a high proportion of repetitive sequences interspersed with single-copy sequences (Pedrosa-Harand et al., 2009).

The organization of the knob satellite *khipu* in *Pv* is similar to the organization of chromosome termini in rice (Ohmido et al., 2001), tomato (*Solanum lycopersicum*; Zhong et al., 1998), and maize (Lamb et al., 2007), where a satellite sequence is also positioned subterminally to telomeres. As previously suggested (Lamb et al., 2007), this could reflect common requirements for chromosome structure at subterminal regions and strongly suggests a functional role for such satellites (Lamb et al., 2007). In addition, we have demonstrated that *khipu* tandem repeats are components of the two knobs located at the B4 *R* gene cluster and are also tightly interleaved between B4-CNL sequences in the B4-CNL subclusters. Altogether, these data strongly suggest a link between *khipu* and the evolution of the B4 *R* gene cluster, arguing that this satellite DNA plays an important evolutionary role and is not merely junk DNA. Data mining of bean BAC end sequence confirmed that *khipu* is abundant in the bean genome, since it is present on 888 (~1%) out of 89,017 available bean BAC end sequence (Schlueter et al., 2008). However, our FISH results indicate that *khipu* is probably not the only component of *Pv* terminal knobs, since five terminal knobs exhibit no *khipu* hybridization signal. Because BAT93 has three terminally located 45S rDNA clusters (Pedrosa-Harand et al., 2006), at least two out of these five terminal knobs probably have other satellite sequences. Alternatively, since satellite DNA is considered a fast-evolving sequence, a *khipu*-related satellite may be present at these two apparently *khipu*-less terminal knobs but too divergent to be detected by our *khipu* probe. The high level of divergence observed between several *khipu* units from the B4 locus, with only 40% nucleotide identity (Supplemental Fig. S6), is in favor of this latter hypothesis.

Given that the CNL ectopic recombination event is shared by *Pv* and *Gm*, a crucial question is why the B4-CNL expanded dramatically in *Pv* but not in *Gm*. Our analysis suggests that subtelomeric location, per se, is not sufficient to explain the CNL proliferation, since no amplification is observed for the subtelomeric CNL on *Gm19*. A first hypothesis is related to the fact that after *Pv* and *Gm* divergence, *Gm* has undergone a whole genome duplication (Shoemaker et al., 2006). By duplicating its whole genome, *Gm* also duplicated its Co-2-CNL orthologs, located at the *Rpg1* cluster (Innes et al., 2008). Thus, CNL proliferation in the syntenic region of the B4 locus in *Gm* (on *Gm19* and *Gm5*) could have been unnecessary and/or could have been counterselected following whole genome duplication because of fitness costs associated with the maintenance of large numbers of CNL (Tian et al., 2003; Nobuta et al., 2005). A second hypothesis is that in the case of the *Pv* B4 locus, *khipu* tandem repeats provide two leads to explain the important B4-CNL proliferation. First, the identification of *khipu* tandem repeats tightly interleaved between B4-CNL sequences (Supplemental Fig. S1) strongly suggests that amplification of the B4-CNL could have been promoted by unequal crossing over involving *khipu* tandem repeats. This is reminiscent of results from flax (*Linum usitatissimum*), where repeat sequences have been identified at the complex *M* resistance locus but not at the simple *L* locus (Ellis et al., 1995). Theoretically, recombination by unequal crossover would lead to tandem arrays of CNL and of *khipu* blocks within a subcluster. This holds true for all of them with the exception of *khipu* tandem repeats C (subcluster C) and K (subcluster A), which are in opposite orientation relative to the others, suggesting the involvement of additional unknown molecular mechanisms for *khipu* amplification. Second, an alternative but not mutually exclusive explanation is that large numbers of *khipu* tandem repeats may have favored the formation of a peculiar genomic environment through their organization into two knobs present at the end of *Pv4*. Indeed, transcripts of satellite sequences are known to initialize heterochromatin formation through an RNA interference mechanism (Pontier et al., 2005; Kloc and Martienssen, 2008). The B4-CNL are located in the euchromatin/heterochromatin junction on both sides of the minor knob. Interestingly, other large *R* gene clusters, such as the tomato *Mi* cluster (Zhong et al., 1999), are also located at the euchromatin/heterochromatin junction. Because epigenetic silencing of genes in proximity to heterochromatic regions is a known process (Talbert and Henikoff, 2006), we hypothesize that this peculiar genomic environment may favor the proliferation of large *R* gene clusters due to some form of silencing (Lippman et al., 2004; Yi and Richards, 2007; Geffroy et al., 2009). Indeed, the importance of RNA silencing to maintain low levels of expression at plant *R* gene clusters was identified by Yi and Richards (2007) for the Arabidopsis *RPP5* locus.

MATERIALS AND METHODS

Plant Material

Seeds of wild beans (*Phaseolus vulgaris*, *P. glabellus*, *P. oligospermus*, *P. pedicellatus*, *P. filiformis*, *P. parvifolius*, *P. dumosus*, *P. leptostachyus*, *P. lunatus*, and *P. microcarpus*) as well as *Centrosema macrocarpum*, *Vigna luteola*, *V. vexillata*, *Teramnus uncinatus*, and *Macroptilium atropurpureum* used in this study were obtained from the Centro Internacional de Agricultura Tropical (Colombia). Seventy-seven F11 recombinant inbred lines derived from a cross between the cultivated Mesoamerican BAT93 genotype and the cultivated Andean JaloEEP558 genotype were used to map Pv BAC clone 68o6 on the integrated linkage map of common bean (Freyre et al., 1998).

BAC Sequencing and Annotation

In a previous study, 73 positive BAC clones corresponding to the B4 *R* gene cluster were organized into six contigs (David et al., 2008). We sequenced six BAC clones representing the minimum tilling path. Shotgun DNA sequencing of the *P. vulgaris* BAC clones and subsequent assembly of the sequence was performed by Lark Technologies and by Genome Express. Gene prediction was done using a combination of gene-finding programs and sequence homology with known genes and proteins. The two ab initio gene prediction programs FGENESH (Burset and Guigo, 1996) and GeneMarkhmm (Lukashin and Borodovsky, 1998) were used. BLAST (Altschul et al., 1997) analyses against the GenBank nonredundant database and all the *Phaseolus* ESTs available at GenBank were performed (Ramirez et al., 2005). Repetitive

elements and transposon sequences were identified using RepeatMasker (http://repeatmasker.org) and The Institute for Genomic Research Arabidopsis repetitive element database. The criteria used to define a gene when there was no EST support were, first, a match to a sequence in a protein database using BLASTX and, second, prediction as a gene by the two prediction programs. These criteria are used after excluding identified transposable elements. The *khipu* satellite DNA was recovered with MEME software (Bailey and Elkan, 1994). The *khipu* repeat units were aligned using ClustalW (Thompson et al., 1994). The alignment was used as input to the profile-building program hmmbuild, which is a program in the HMMER package (http://hmmer.janelia.org; Eddy, 1998). The resulting profile was used for searching for additional *khipu* units in the 650-kb sequence using hmmsearch (another program in the HMMER package). All this information was imported into the annotation platform Artemis for further manual analysis (Rutherford et al., 2000). Sequences were considered as pseudogenes when premature stop codons or frameshifts were identified. An asterisk has been added at the end of CNL sequence names corresponding to pseudogenes.

Sequence Alignment and Computational Analyses

MAFFT version 6 was employed for multiple nucleic acid and amino acid alignments using the L-INS-i strategy (Katoh et al., 2005) with default parameters, and GENEDOC (http://www.psc.edu/biomed/genedoc) was used for manual adjustment of subsequent alignments.

For CNL phylogenetic analyses, two alignments were carried out based on (1) 26 nucleotide BAT93 B4-CNL sequences (Fig. 2) and (2) 133 NL amino acid sequences from legumes, Arabidopsis (*Arabidopsis thaliana*), and poplar (*Populus trichocarpa*; Fig. 3). The first alignment (Fig. 2) was generated using the entire NBS motif (from P-loop to MHDV) for the 26 BAT93 B4-CNL nucleotide sequences. The soybean (*Glycine max*) *Rpg1-b* CNL sequence was used as an outgroup sequence (Ashfield et al., 2004). The second alignment (Fig. 3) of the 133 amino acid NL sequences is based on the NBS amino acid sequence region between the P-loop domain and the GLPL domain. Twenty-six BAT93 B4-CNL were identified in this study in the approximately 650-kb region. Fourteen out of 17 previously identified BAT93 B4-CNL were recovered in this study (Ferrier Cana et al., 2003; Geffroy et al., 2009). The three sequences not recovered in this study are CNL-BA8, CNL-BA11 (Ferrier Cana et al., 2003), and CNL-B19 (Geffroy et al., 2009). For this second alignment, out of the 29 BAT93 B4-CNL, we excluded the B4-CNL corresponding to pseudogenes and used the amino acid sequences of 20 BAT93 *Pv* B4-CNL and added seven JaloEEP558 *Pv* B4-CNL. Additionally, the following *Lotus japonicus* and *Medicago truncatula* genes were added, following the description and nomenclature of Ameline-Torregrosa et al. (2008): *Lj3fg26877*, *Lj3fg26933*, *Mt8g1223* (AC153125_16.5), *Mt8g1225* (AC153125_27.5), *Mt2g1873* (AC148761_18.3), *Mt2g1878* (AC148761_27.3), and *Mt2g1862* (AC148761_2.3). *Mt8g1223* and *Mt8g1225* are locally duplicated genes on chromosome 8, and *Mt2g1873*, *Mt2g1878*, and *Mt2g1862* are locally duplicated on chromosome 2 (Ameline-Torregrosa et al., 2008; Innes et al., 2008). Separation of these chromosomal regions on *Mt2* and *Mt8* may date to a whole genome duplication that occurred early in legume evolution (Shoemaker et al., 2006). Also included were known orthologs of *Mt8g1223* and *Mt8g1225* sequences in soybean (43 sequences) and common bean (20 sequences; Innes et al., 2008), four additional members of the *Gm Rpg1* cluster described by Ashfield et al. (2004) and Hayes et al. (2004), and eight additional *Mt* sequences from the CNL-2 to CNL-5 clades described by Ameline-Torregrosa et al. (2008). Eight homologs from other sequenced genomes (Arabidopsis and poplar) described by Ameline-Torregrosa et al. (2008) were also included. Searches of NL (CNL and TNL) sequences from *Pv* in the nr database with both BLASTX (cutoff value of 10^{-5}) and TBLASTX (cutoff value of 10^{-3}) led to the identification of 16 additional amino acid *Pv* NL sequences (including 11 TNL sequences). The NBS amino acid sequences (from the P-loop to the GLPL domain) of these 11 *Pv* TNL identified by TBLASTX and BLASTX against the nr database were used as an outgroup to root the tree. All the sequences used to generate both alignments are reported in Supplemental Table S5.

Graphical representation of the *khipu* alignment generated using the 92 *khipu* unit nucleic acid sequences identified in the B4 *R* gene cluster was visualized using the WebLogo server (http://weblogo.berkeley.edu; Crooks et al., 2004). Nucleotide identities were obtained using bl2seq (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). Assessment of positive selection was performed as described by Geffroy et al. (2009) using the program CODEML from the PAML package (Yang, 1997).

Phylogenetic Analyses

These two NL alignments were both analyzed using Bayesian methods and maximum likelihood (ML) as implemented in MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001) and PhyML (Guindon and Gascuel, 2003), respectively. MrBayes analyses were run with the nucleotide model 4by4 (for nucleic acid alignment) or a fixed Jones amino acid substitution model (for amino acid alignment), following a γ distribution. We used the Metropolis-coupled Markov Chain Monte Carlo method to make two runs of four chains. We ran 10 million generations, sampling every 100 generations. A consensus tree with branch lengths and posterior estimates of branch probabilities was generated with the "sumt" command of MrBayes and a "burnin" parameter of 50,000. ML analysis was carried out with the GTR model for nucleic acid alignment and the JTT model for amino acid alignment. Relative supports of both analyses were assessed with 1,000 bootstrap replicates. Resulting phylogenetic trees were displayed with the MEGA software (Tamura et al., 2007). The resulting ML and MrBayes trees have a similar topology, and only Bayesian trees are presented.

Microsynteny Analysis

Putative gene homologs of the *Pv* B4 locus 410-kb contig sequence were identified by TBLASTN searches in *Mt*, *Lj*, and *Gm* genomes (cutoff value of $1e^{-5}$) against all *Mt* (www.medicago.org), *Lj* (http://www.kazusa.or.jp/lotus/), and *Gm* (http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01/) BAC sequences available on June 26, 2007, June 4, 2008, and January 7, 2009, respectively. Sequence annotation of the putative syntenic BAC clones from *Mt*, *Lj*, and *Gm* was performed as described for the *Pv* B4 650-kb sequence. In order to confirm that these BACs were syntenic with the *Pv* B4 locus, and to identify syntenic genes, putative *Pv* B4 versus *Mt*, *Lj*, and *Gm* protein comparisons were performed (Supplemental Table S6). Genetic and physical map information was further searched for *Mt* (www.medicago.org), *Lj* (http://www.kazusa.or.jp/lotus/), and *Gm* (http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01/) positive syntenic BAC clones.

Molecular Analysis

DNA extraction, Southern-blot hybridization, and PCR experiments were carried out as described by Geffroy et al. (1998). Oligonucleotide primers SatF1 (5′-GAGGATTATCACTGCCTGAC-3′) and SatR1 (5′-CTTCAGCCAA-CTAGTTATGTAC-3′) were designed to amplify *khipu* tandem repeat using BAT93 BAC clone 75-H11 as a template. The 3,978-bp amplification product obtained with SatF1 and SatR1 is referred to as SatF1R1. SatF1R1 sequence was used as a probe in hybridization experiments.

Linkage Analysis

A PCR-based approach was used to map *Pv* BAC clone 68o6, syntenic to the soybean *Rpg1* locus described by Innes et al. (2008), on the common bean integrated genetic map (Freyre et al., 1998). Specific oligonucleotide primers 68o6-f1.1 (5′-GCCTGATGCCAACAAGTACA-3′) and 68o6-f1.2 (5′-AAGG-ATGAACTGCCTCATGC-3′) derived from BAC clone 68o6 (accession no. CU914569) were used to amplify genomic DNA of the 77 BAT93 × JaloEEP558 F9 recombinant inbred lines. Presence/absence polymorphisms were scored. The MAPMAKER software version 3.0 was used (Lander et al., 1987). Linkage groups were established with a log of the odds threshold of 3.0 and a recombination fraction of 0.3. Marker order was estimated with a log of the odds threshold of 2.0 based on multipoint "compare," "order," and "ripple" analyses. Distances between markers were presented in cM (Kosambi, 1944).

FISH Mapping

Two BAT93 clones were used as probes: the ER-N1 BAC clone (David et al., 2008), harboring *Pv* B4-CNL sequences but no *khipu* tandem repeat (position on the *Pv* 410-kb contig, 6,978–112,554 bp), and the shotgun plasmid subclone 1H04, containing five *khipu* units (position on the *Pv* 410-kb contig, 390,736–388,326 bp). Probes were labeled by nick translation (Invitrogen) with Cy3-dUTP (General Electric).

Mitotic chromosome preparation was as described by Pedrosa-Harand et al. (2006). Meiotic chromosomes were prepared from young flower buds fixed in ethanol:acetic acid (3:1, v/v). Buds were macerated in 2% cellulase/

2% pectolyase/2% citohelicase in 0.01 M citric acid-sodium citrate buffer, pH 4.8, for 3 h at 37°C, incubated in 60% acetic acid up to 2 h, and squashed after removal of petals and sepals. Slide selection and pretreatment were as described by Pedrosa et al. (2001). Chromosome and probe denaturation and posthybridization washes were performed according to Heslop-Harrison et al. (1991), with modifications described by Pedrosa-Harand et al. (2006), except that meiotic preparations were denatured at 73°C for 3 min. Reprobing of slides was performed according to Heslop-Harrison et al. (1992).

Images were captured on a Leica DMLB microscope equipped with a CCD Cohu camera and combined using the QFISH software (Leica). Digital images were imported into Adobe Photoshop version 8 for final processing.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers FJ817291 (*Pv* 410-kb contig), FJ817289 (48-B10 BAC clone), and FJ817290 (FZ-E9 BAC clone).

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Simplified representation of the 650-kb sequence corresponding to the *Pv* B4 resistance locus showing details concerning *khipu* blocks.

Supplemental Figure S2. Sites identified as being under diversifying selection on the amino acid sequence for the predicted full-length transcript of B4-CNL-B7.

Supplemental Figure S3. Sequence comparison between the *Pv* BAT93 B4 410-kb contig and syntenic regions identified in the *Gm* genome.

Supplemental Figure S4. Amino acid alignment of CNL-B17 and the two truncated *Gm* B4-CNL identified on the soybean B4 syntenic region.

Supplemental Figure S5. WebLogo representation of the consensus sequence derived from multiple alignment of the 92 *Pv* B4-*khipu* units.

Supplemental Figure S6. Frequencies of nucleotide identities from pairwise comparisons between 92 B4-*khipu* units in the *Pv* B4 650-kb region.

Supplemental Figure S7. Occurrence and distribution of the *khipu* satellite DNA and the B4-CNL sequences in various legume species.

Supplemental Figure S8. Hypothetical model of evolution of the B4 and Co-2 resistance loci from the *Mt*, *Lj*, *Gm*, and *Pv* common ancestor to the current state.

Supplemental Table S1. Characteristics of the predicted genes identified in the 650-kb *Pv* BAT93 B4 *R* gene locus sequence.

Supplemental Table S2. *Phaseolus* EST BLASTN hits on the 650 kb spanning the *Pv* BAT93 B4 *R* gene locus.

Supplemental Table S3. Distribution of transposable elements identified in the 650 kb of the *Pv* BAT93 B4 *R* gene locus sequence.

Supplemental Table S4. Percentage of nucleotide sequence identity between the 26 BAT93 B4-CNL.

Supplemental Table S5. Origin of NBS-LRR sequences used in this study.

Supplemental Table S6. Percentage of amino acid similarity between *Pv*, *Lj*, *Mt*, and *Gm* genes identified in segments presenting conserved synteny.

ACKNOWLEDGMENTS

We thank the RPG1 consortium (Innes et al., 2008) for sharing unpublished data and Paul Gepts and Jim Kami for providing the BAT93 BAC clones. We thank Jeff J. Doyle for helpful discussion.

Received May 26, 2009; accepted September 17, 2009; published September 23, 2009.

LITERATURE CITED

Adam-Blondon AF, Sévignac M, Bannerot H, Dron M (1994) SCAR, RAPD and RFLP markers linked to a dominant gene (*Are*) conferring resistance to anthracnose in common bean. Theor Appl Genet **88:** 865–870

Alexandrov I, Kazakov A, Tumeneva I, Shepelev V, Yurov Y (2001) Alpha-satellite DNA of primates: old and new families. Chromosoma **110:** 253–266

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389–3402

Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu HY, Roe B, Young ND, Cannon SB (2008) Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiol **146:** 5–21

Ammiraju JSS, Lu F, Sanyal A, Yu Y, Song X, Jiang N, Pontaroli AC, Rambo T, Currie J, Collura K, et al (2008) Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. Plant Cell **20:** 3191–3209

Ananiev EV, Phillips RL, Rines HW (1998) A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? Proc Natl Acad Sci USA **95:** 10785–10790

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815

Ashfield T, Ong LE, Nobuta K, Schneider CM, Innes RW (2004) Convergent evolution of disease resistance gene specificity in two flowering plant families. Plant Cell **16:** 309–318

Bai JF, Pennill LA, Ning JC, Lee SW, Ramalingam J, Webb CA, Zhao BY, Sun Q, Nelson JC, Leach JE, et al (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res **12:** 1871–1884

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol **2:** 28–36

Baumgarten A, Cannon S, Spangler R, May G (2003) Genome-level evolution of resistance genes in *Arabidopsis thaliana*. Genetics **165:** 309–319

Bennett MD, Leitch IJ (1995) Nuclear-DNA amounts in angiosperms. Ann Bot (Lond) **76:** 113–176

Bennett MD, Leitch IJ (1997) Nuclear DNA amounts in angiosperms: 583 new estimates. Ann Bot (Lond) **80:** 169–196

Bennett MD, Smith JB (1991) Nuclear-DNA amounts in angiosperms. Philos Trans R Soc Lond B **334:** 309–345

Bennetzen JL, Ma JX (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol **6:** 128–133

Bertioli DJ, Moretzsohn MC, Madsen LH, Sandal N, Leal-Bertioli SCM, Guimaraes PM, Hougaard BK, Fredslund J, Schauser L, Nielsen AM, et al (2009) An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. BMC Genomics **10:** 11

Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.): model food legumes. Plant Soil **252:** 55–128

Burset M, Guigo R (1996) Evaluation of gene structure prediction programs. Genomics **34:** 353–367

Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang XH, Mudge J, Vasdewani J, Scheix T, et al (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc Natl Acad Sci USA **103:** 14959–14964

Chacon MI, Pickersgill B, Debouck DG, Arias JS (2007) Phylogeographic analysis of the chloroplast DNA variation in wild common bean (*Phaseolus vulgaris* L.) in the Americas. Plant Syst Evol **266:** 175–195

Charron MJ, Read E, Haut SR, Michels CA (1989) Molecular evolution of the telomere-associated *MAL* loci of *Saccharomyces*. Genetics **122:** 307–316

Choi HK, Mun JH, Kim DJ, Zhu HY, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, et al (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA **101:** 15289–15294

Creusot F, Macadré C, Ferrier Cana E, Riou C, Geffroy V, Sévignac M, Dron M, Langin T (1999) Cloning and molecular characterization of three members of the NBS-LRR subfamily located in the vicinity of the *Co-2* locus for anthracnose resistance in *Phaseolus vulgaris*. Genome **42:** 254–264

Cronk Q, Ojeda I, Pennington RT (2006) Legume comparative genomics: progress in phylogenetics and phylogenomics. Curr Opin Plant Biol **9:** 99–103

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res **14:** 1188–1190

Dangl JL, McDowell JM (2006) Two modes of pathogen recognition by plants. Proc Natl Acad Sci USA **103:** 8575–8576

David P, Sévignac M, Thareau V, Catillon Y, Kami J, Gepts P, Langin T, Geffroy V (2008) BAC end sequences corresponding to the B4 resistance gene cluster in common bean: a resource for markers and synteny analyses. Mol Genet Genomics **280:** 521–533

Delgado-Salinas A, Bibler R, Lavin M (2006) Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. Syst Bot **31:** 779–791

Eddy SR (1998) Hidden Markov models and genome sequence analysis. FASEB J **12:** A1327

Ellis JG, Lawrence EJ, Finnegan EJ, Anderson PA (1995) Contrasting complexity of two rust resistance loci in flax. Proc Natl Acad Sci USA **92:** 4185–4188

Fan C, Zhang Y, Yu Y, Rounsley S, Long M, Wing RA (2008) The subtelomere of *Oryza sativa* chromosome 3 short arm as a hot bed of new gene origination in rice. Mol Plant **1:** 839–850

Ferrier Cana E, Geffroy V, Macadre C, Creusot F, Imbert Bollore P, Sevignac M, Langin T (2003) Characterization of expressed NBS-LRR resistance gene candidates from common bean. Theor Appl Genet **106:** 251–261

Ferrier Cana E, Macadre C, Sevignac M, David P, Langin T, Geffroy V (2005) Distinct post-transcriptional modifications result into seven alternative transcripts of the CC-NBS-LRR gene JA1tr of *Phaseolus vulgaris*. Theor Appl Genet **110:** 895–905

Fransz PF, Armstrong S, de Jong JH, Parnell LD, van Drunen G, Dean C, Zabel P, Bisseling T, Jones GH (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. Cell **100:** 367–376

Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D (2008) Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. Genome Res **18:** 1924–1937

Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, Nehrbass U, Wellems TE, Scherf A (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. Nature **407:** 1018–1022

Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, Rojas-Meza AP, Mancio-Silva L, Leal-Silvestre RJ, Gontijo AM, Shorte S, et al (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell **121:** 25–36

Freyre R, Skroch P, Geffroy V, Adam-Blondon AF, Shirmohamadali A, Johnson WC, Llaca V, Nodari RO, Pereira PA, Tsai SM, et al (1998) Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. Theor Appl Genet **97:** 847–856

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK (2007) Opinion. Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet **8:** 77–84

Geffroy V, Creusot F, Falquet J, Sevignac M, Adam-Blondon AF, Bannerot H, Gepts P, Dron M (1998) A family of LRR sequences in the vicinity of the *Co-2* locus for anthracnose resistance in *Phaseolus vulgaris* and its potential use in marker-assisted selection. Theor Appl Genet **96:** 494–502

Geffroy V, Macadré C, David P, Pedrosa-Harand A, Sevignac M, Dauga C, Langin T (2009) Molecular analysis of a large subtelomeric nucleotide-binding-site-leucine-rich-repeat family in two representative genotypes of the major gene pools of *Phaseolus vulgaris*. Genetics **181:** 405–419

Geffroy V, Sevignac M, Billant P, Dron M, Langin T (2008) Resistance to *Colletotrichum lindemuthianum* in *Phaseolus vulgaris*: a case study for mapping two independent genes. Theor Appl Genet **116:** 407–415

Geffroy V, Sevignac M, De Oliveira JCF, Fouilloux G, Skroch P, Thoquet P, Gepts P, Langin T, Dron M (2000) Inheritance of partial resistance against *Colletotrichum lindemuthianum* in *Phaseolus vulgaris* and co-localization of quantitative trait loci with genes involved in specific resistance. Mol Plant Microbe Interact **13:** 287–296

Geffroy V, Sicard D, de Oliveira JCF, Sevignac M, Cohen S, Gepts P, Neema C, Langin T, Dron M (1999) Identification of an ancestral resistance gene cluster involved in the coevolution process between *Phaseolus vulgaris* and its fungal pathogen *Colletotrichum lindemuthianum*. Mol Plant Microbe Interact **12:** 774–784

Gonzalez-Garcia M, Gonzalez-Sanchez M, Puertas MJ (2006) The high variability of subtelomeric heterochromatin and connections between nonhomologous chromosomes, suggest frequent ectopic recombination in rye meiocytes. Cytogenet Genome Res **115:** 179–185

Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. Plant Physiol **131:** 872–877

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:** 696–704

Hayes AJ, Jeong SC, Gore MA, Yu YG, Buss GR, Tolin SA, Maroof MAS (2004) Recombination within a nucleotide-binding-site/leucine-rich-repeat gene cluster produces new variants conditioning resistance to soybean mosaic virus in soybeans. Genetics **166:** 493–503

Heacock M, Spangler E, Riha K, Puizina J, Shippen DE (2004) Molecular analysis of telomere fusions in *Arabidopsis*: multiple pathways for chromosome end-joining. EMBO J **23:** 2304–2313

Heslop-Harrison JS, Harrison GE, Leitch IJ (1992) Reprobing of DNA: DNA *in situ* hybridization preparations. Trends Genet **8:** 372–373

Heslop-Harrison JS, Schwarzacher T, Anamthawat-Jónsson K, Leitch AR, Shi M, Leitch IJ (1991) *In situ* hybridization with automated chromosome denaturation. Technique **3:** 109–115

Hougaard BK, Madsen LH, Sandal N, Moretzsohn MD, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S, et al (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. Genetics **179:** 2299–2312

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:** 754–755

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NWG, Couloux A, Dalwani A, Denny R, et al (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol **148:** 1740–1759

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res **33:** 511–518

Kawabe A, Charlesworth D (2007) Patterns of DNA variation among three centromere satellite families in *Arabidopsis halleri* and *A. lyrata*. J Mol Evol **64:** 237–247

Kloc A, Martienssen R (2008) RNAi, heterochromatin and the cell cycle. Trends Genet **24:** 511–517

Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F (2008) Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. Plant Mol Biol **66:** 619–636

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen **12:** 172–175

Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell **16:** 2870–2894

Kulikova O, Gualtieri G, Geurts R, Kim DJ, Cook D, Huguet T, de Jong JH, Fransz PF, Bisseling T (2001) Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. Plant J **27:** 49–58

Kuo HF, Olsen KM, Richards EJ (2006) Natural variation in a subtelomeric region of Arabidopsis: implications for the genomic dynamics of a chromosome end. Genetics **173:** 401–417

Lamb JC, Meyer JM, Corcoran B, Kato A, Han FP, Birchler JA (2007) Distinct chromosomal distributions of highly repetitive sequences in maize. Chromosome Res **15:** 33–49

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol **54:** 575–594

Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet **20:** 116–122

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci USA **95:** 370–375

Linardopoulou EV, Williams EM, Fan YX, Friedman C, Young JM, Trask BJ (2005) Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. Nature **437:** 94–100

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, et al (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature **430:** 471–476

Lopez CE, Acosta IF, Jara C, Pedraza F, Gaitan-Solis E, Gallego G, Beebe S, Tohme J (2003) Identifying resistance gene analogs associated with resistances to different pathogens in common bean. Phytopathology **93:** 88–95

Louis EJ (1995) The chromosome ends of *Saccharomyces cerevisiae*. Yeast **11:** 1553–1573

Lu F, Ammiraju JSS, Sanyal A, Zhang S, Song R, Chen J, Li G, Sui Y, Cheng Z, Costa de Oliveira A, et al (2009) Comparative sequence analysis of the MONOCULM1 orthologous regions in 14 *Oryza* genomes. Proc Natl Acad Sci USA **106:** 2071–2076

Lukashin AV, Borodovsky M (1998) GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res **26:** 1107–1115

Ma JX, Jackson SA (2006) Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. Genome Res **16:** 251–259

McDowell JM, Simon SA (2006) Recent insights into *R* gene evolution. Mol Plant Pathol **7:** 437–448

McDowell JM, Simon SA (2008) Molecular diversity at the plant-pathogen interface. Dev Comp Immunol **32:** 736–744

Mefford HC, Trask BJ (2002) The complex structure and dynamic evolution of human subtelomeres. Nat Rev Genet **3:** 91–102

Melotto M, Coelho MF, Pedrosa Harand A, Kelly JD, Camargo LEA (2004) The anthracnose resistance locus *Co-4* of common bean is located on chromosome 3 and contains putative disease resistance-related genes. Theor Appl Genet **109:** 690–699

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell **10:** 1817–1832

Meyers BC, Kozik A, Griego A, Kuang HH, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell **15:** 809–834

Miklas PN, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. Euphytica **147:** 105–131

Ming R, Hou SB, Feng Y, Yu QY, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KLT, et al (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature **452:** 991–996

Nobuta K, Ashfield T, Kim S, Innes RW (2005) Diversification of non-TIR class NB-LRR genes in relation to whole-genome duplication events in *Arabidopsis*. Mol Plant Microbe Interact **18:** 103–109

Ohmido N, Kijima K, Ashikawa I, de Jong JH, Fukui K (2001) Visualization of the terminal structure of rice chromosomes 6 and 12 with multicolor FISH to chromosomes and extended DNA fibers. Plant Mol Biol **47:** 413–421

Ohno S (1972) So much "junk" DNA in our genome. Brookhaven Symp Biol **23:** 366–370

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature **284:** 604–607

Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ (1981) Highly repeated DNA-sequence limited to knob heterochromatin in maize. Proc Natl Acad Sci USA **78:** 4490–4494

Pedrosa A, Jantsch MF, Moscone EA, Ambros PF, Schweizer D (2001) Characterisation of pericentromeric and sticky intercalary heterochromatin in *Ornithogalum longibracteatum* (Hyacinthaceae). Chromosoma **110:** 203–213

Pedrosa A, Sandal N, Stougaard J, Schweizer D, Bachmair A (2002) Chromosomal map of the model legume *Lotus japonicus*. Genetics **161:** 1661–1672

Pedrosa-Harand A, de Almeida CCS, Mosiolek M, Blair M, Schweizer D, Guerra M (2006) Extensive ribosomal DNA amplification during Andean common bean (*Phaseolus vulgaris* L.) evolution. Theor Appl Genet **112:** 924–933

Pedrosa-Harand A, Kami J, Gepts P, Geffroy V, Schweizer D (2009) Cytogenetic mapping of common bean chromosome reveals a less compartmentalized small-genome plant species. Chromosome Res **17:** 405–417

Plohl M, Luchetti A, Mestrovic N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene **409:** 72–82

Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, Lagrange T (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. Genes Dev **19:** 2030–2040

Ramirez M, Graham MA, Blanco-Lopez L, Silvente S, Medrano-Soto A, Blair MW, Hernandez G, Vance CP, Lara M (2005) Sequencing and analysis of common bean ESTs: building a foundation for functional genomics. Plant Physiol **137:** 1211–1227

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. Bioinformatics **16:** 944–945

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, et al (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res **15:** 227–239

Schlueter JA, Goicoechea JL, Collura K, Gill N, Lin JY, Yu Y, Kudrna D, Zuccolo A, Vallejos CE, Munoz-Torres M, et al (2008) BAC-end sequence analysis and a draft physical map of the common bean (*Phaseolus vulgaris* L.) genome. Trop Plant Biol **1:** 10–48

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol **9:** 104–109

Smith SM, Pryor AJ, Hulbert SH (2004) Allelic and haplotypic diversity at the *Rp1* rust resistance locus of maize. Genetics **167:** 1939–1947

Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, Ravetch JA, Wellems TE (1995) The large diverse gene family var encodes protein involved in cytoadherance and antigenic variation of *Plasmodium falciparum* infected erythrocytes. Cell **82:** 89–100

Talbert PB, Henikoff S (2006) Spreading of silent chromatin: inaction at a distance. Nat Rev Genet **7:** 793–803

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24:** 1596–1599

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:** 4673–4680

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. Nature **423:** 74–77

Tsubokura Y, Onda R, Sato S, Xia ZJ, Hayashi M, Fukushima Y, Tabata S, Harada K (2008) Characterization of soybean genome based on synteny analysis with Lotus japonicus. Breed Sci **58:** 157–167

Turakainen H, Naumov G, Naumova E, Korhola M (1993) Physical mapping of the *MEL* gene family in *Saccharomyces cerevisiae*. Curr Genet **24:** 461–464

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596–1604

Vallejos CE, Astua-Monge G, Jones V, Plyler TR, Sakiyama NS, Mackenzie SA (2006) Genetic and molecular characterization of the *I* locus of *Phaseolus vulgaris*. Genetics **172:** 1229–1242

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, et al (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One **2:** e1326

Vershinin AV, Alkhimova EG, Heslop-Harrison JS (1996) Molecular diversification of tandemly organized DNA sequences and heterochromatic chromosome regions in some Triticeae species. Chromosome Res **4:** 517–525

Waye JS, Willard HF (1989) Concerted evolution of alpha satellite DNA: evidence for species specificity and a general lack of sequence conservation among alphoid sequences of higher primates. Chromosoma **98:** 273–279

Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci **13:** 555–556

Yi H, Richards EJ (2007) A cluster of disease resistance genes in *Arabidopsis* is coordinately regulated by transcriptional activation and RNA silencing. Plant Cell **19:** 2929–2939

Young ND, Cannon SB, Sato S, Kim D, Cook DR, Town CD, Roe BA, Tabata S (2005) Sequencing the gene spaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiol **137:** 1174–1181

Zhong XB, Bodeau J, Fransz PF, Williamson VM, van Kammen A, de Jong JH, Zabel P (1999) FISH to meiotic pachytene chromosomes of tomato locates the root knot nematode resistance gene *Mi-1* and the acid phosphatase gene *Aps-1* near the junction of euchromatin and pericentromeric heterochromatin of chromosome arms 6S and 6L, respectively. Theor Appl Genet **98:** 365–370

Zhong XB, Fransz PF, Wennekes-van Eden J, Ramanna MS, van Kammen A, Zabel P, de Jong JH (1998) FISH studies reveal the molecular and chromosomal organization of individual telomere domains in tomato. Plant J **13:** 507–517

Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genomics **271:** 402–415

Zhu HY, Choi HK, Cook DR, Shoemaker RC (2005) Bridging model and crop legumes through comparative genomics. Plant Physiol **137:** 1189–1196